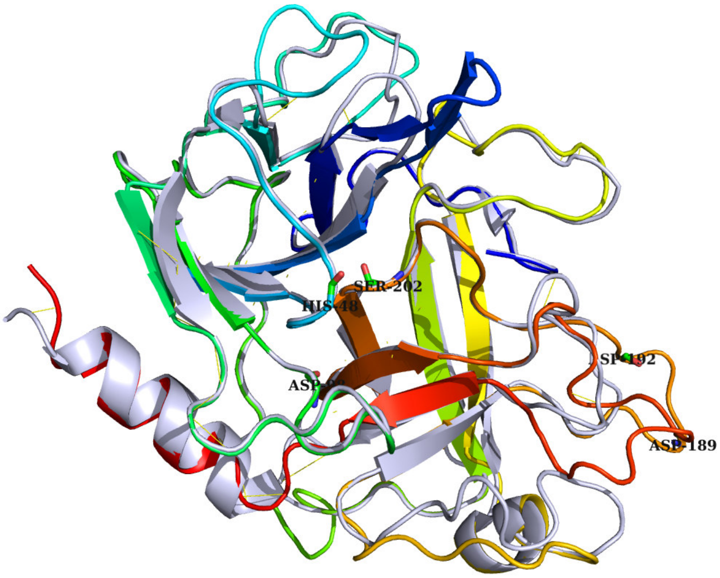
Identification, Cloning and Expressions of Proteases from a Cold Adapted Organism *Aliivibrio salmonicida*



KJE-3900

Arjumand Ather

Master thesis in structural biology
Faculty of science
University of Tromsø
July 2009

Cover page illustration: Model of TVS4041 (coloured chainbow), superimpose with template, human granzyme A ‘**1OP8’** (coloured gray).

Dedicated to my parents.

'For what I am'.

Are their contributions…

Genetical and moral.

Dedicated to my children.'

For what they are'.

Have some part of mine…

After me.

Genetical and moral.

Acknowledgement

I am earnestly grateful to my supervisor and mentor Arne O. smalas for believing in me and giving me moral support in the extremeness of research from office desk to, the laboratory bench. Due to his innovative and creative ideas I can be able to perform the assigned task in such a comprehensive way. I am obliged to thanks Nils Peder Willasen, for giving me opportunity to work in his group on Aliivibrio salmonicida (Strain LFI1238) genome project, due to his kind appreciation this thesis is now in this present form. I am also grateful to thesis work advisors, Marit Sjo Lorentzen, Ronny Helland for their sincerest wishes, timely help, appreciation of my unskilled ideas and constant supervision throughout my research work. Indeed I also appreciate the kind and experienced advises from Bjørn Altamarek, Atle Larsen and Eve Gry, due to their advises I can be able to find the correct way to obtain the assigned task. I will not hesitate to mention my colleges due to them I always added up and benefited discussing the problems in the field of structural biology. I also acknowledge Ingebrigt Sylte for his kind permission to use his lab facilities in bioinformatics work.

Last but not least for everybody in this group for their kind smile and warmness that has supported me for performing my job in happy, healthy, homely environment. I acknowledge University of Tromsø for overwhelming me and financially supporting my research in such a worth seen country.

Above in all my primary thanks always reserve to Almighty God who has given me strength to stood up and face all the harshness in the pathways of life and research.

I am indebted to my husband, who has introduced me toward the research field with his unlashed, impressive, ever bright ideas. During this period he was always been the source of constant courage and confidence for me.

Contents

Summary

The work presented in this thesis provides an overview of molecular and structural biology projects and their related problems. The problems arises during these projects are common for most of the target proteins, with varying intensity of severity. The realization and identification of these problems and selection of their correct, best possible solution is an imperative step that might lead to achievement of the goals. This project has started with the survey of a cold adapted bacterial genome for special kind of enzymes 'Proteases' that can hydrolyze other proteins. Working with these kinds of proteins has also increased the set of problems caused by instability due to cold adaptation and autolysis phenomenon related to proteases.

Eight different proteases were chosen as targets, to be cloned, expressed, characterized and structurally resolved. The initial step was the bioinformatics analysis of chosen target that was essentially important to identify the function of domain in any chosen target. This identification helped making the decision to eliminate the unnecessary domains that can influence the overall success of the project. As one of our selected target TVS4041 (trypsin) has suffered the problems in expression, solubility and purification, due to the accumulation of hydrophobic residues in exposed C-terminal tail.

In this research project we have attempted, two type of cloning techniques; Gateway cloning and traditional restriction digestion cloning. Gateway™ cloning technology was proved good for reliability, accuracy and importantly for the facilitation of switch-ability between expression vectors and expression host. This particular system could also try to be conserve, for time and efforts by 'One-Tube BP and LR Gateway™ Reaction'. On the other hand traditional restriction digestion cloning for TVS4041 was also useful for direct one step transformation toward expression host but, with disadvantage of lack of switch-ability for different vectors and expression host.

From the three expressed targets LexA, HslV and TVS4041 two were selected for extended expression trials for increased yield and solubility. Both selected target TVS4041 and LexA, were subjected to two different solubility enhancing techniques, these were varying expression condition and solubility tag attachment.

TVS4041 was facing the problem of insolubility with increasing time of expression. Different expression conditions were tested and resulted in lower pH, low salt and reduced time of expression conditions for highest solubility. It was also suggested that lower pH (6) and absence of salt could be ideal during purification since in these condition, TVS4041 will be relatively inactive. Furthermore it is suggested that detected, insolubility residing C-terminal tail should be clipped and re-clone for prevention of aggregation during purification. In order to lowering the purification efforts and cost, it is suggested that methods should be experimentally evaluated for secretion of this protein into the media.

The experimental work conducted with LexA expression in conjunction to different fusion tags resulted in order of Gb1>NusA>Z=Trx>MBP>6xHis, which was different then the results obtain from other proteins from same organism. These results confirm the complex nature of protein and prove that different protein behave uniquely in response to different fusion tags, irrespective of their belonging with same organism (Braud, Moutiez et al. 2005).

Introductions

1.1.0. Introduction to proteases

Protease, peptidases, proteolytic or peptide bond hydrolytic enzymes are the species of catalytic enzymes that can hydrolyze or break up the peptide bond between two adjacent amino acids.

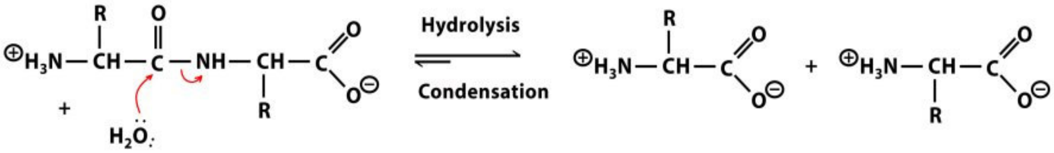


Figure1.1.0: peptide hydrolysis mechanism.

According to Barrett more than 2,000 peptidase species have been recognized until now and they are listed in the MEROPS database (http://merops.sanger.ac.uk/). Among all of the known proteolytic sequences more than two-third of the entire database which present in MEROPS is under the category of unassigned peptidase. This is because they are not sufficiently similar to the holotype of any existing peptide species, but eventually most of them can be put into a newly recognized species of same holotype and specificity (Barrett and Rawlings 2007).

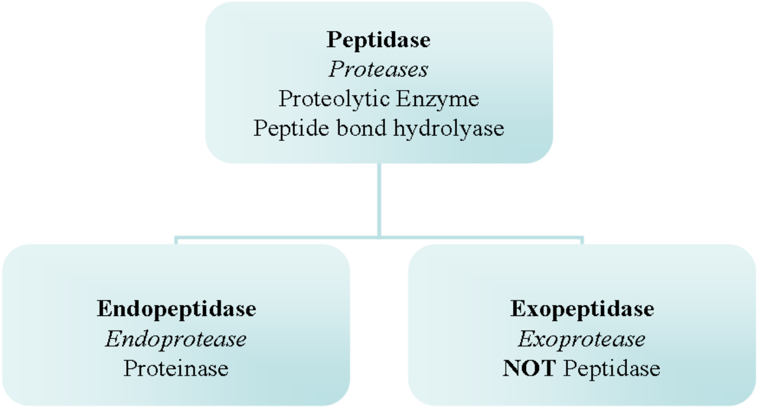
.

1.2.0. Protease and/or Peptidase

The term protease evolved in German physiological chemistry literature during the later part of ninetieth century. The same group has then utilized the term “proteinase” and “peptidase” for the enzyme acting on proteins and peptidase, respectively (Grassmann and Dyckerhoff 1928).

While (Bergmann and Ross 1936) had introduced the term peptides as a general one for any peptide bond hydrolyzing enzyme and further categorized them in endopeptidase and exopeptidase depending on their site of action. That is usually away from the terminus of proteins in case of endopeptidases and to the terminus of peptide in case of exopeptidases.

The latter term was approved by EC's Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (**NC-IUBMB**) as it was further revealed that the tendency of the exopeptidase to act mostly on the peptide is due to the requirement of free terminal group close to scissile bond, that is usually rear in intact proteins. Hence the terminology Oligopeptidase was introduce to define the exopeptidase that act only on peptide (Barrett and McDonald 1985).



(Act away from terminus, inside the proteins and produce relatively large peptides)

(Act near *N*- or *C*- terminus of proteins or peptides and produce small peptide/a.a.)

Figure 1.2.0: Different synonyms of peptidase: The term written in bold are those recommended by IUBMB with the clear logic hence Exopeptidase can not be termed as peptidase as it is also using to describe the entire proteolytic enzymes.

1.3.0. Protease classification

Nomenclature and classification are vitally important for information handling. They allow people to communicate efficiently, with complete understanding of what they are talking about and to store and retrieve information efficiently and unambiguously. A good system must have vast criteria to cover up the entire stream of present data and be able to quickly absorb the coming candidate in correct place. There are three useful orthogonal methods of grouping peptidases:

1. By the chemical mechanism of catalysis.
2. By the position of action, way and mechanism.
3. By molecular structure and homology.

1.3.1. Grouped by the chemical mechanism of catalysis

In 1960 Hartley initiated the categorization of peptide molecules on the bases of their main catalytic residue type, hence a very useful concept emerged that further becames the base of modern peptide classification system. This system brought molecular structures under an umbrella on the bases of catalytic mechanisms, like serine, cysteine, threonine, aspartic, glutamic or metallo "catalytic type" peptidase (Hartley 1960). But most often they do not have any significant homology to each other, and must be categorized further.

1.3.2. Grouped by the position of action, way and mechanism

This classification has further categorized proteases on the bases of, position related selectivity for the hydrolysis in similar group. On this basis, they can be classified into two major and two minor sub classes, namely Endopeptidases (acting away from terminus of larger proteins), Exopeptidase (acting on terminus of polypeptide), Oligopeptidases (acting away from terminus in smaller proteins) and Omega-peptidases (acting on the terminus of proteins) respectively.

The two major classes are then divided in sub-subclasses depending on their site and way of action and some where on the basis of mechanism of action (Barrett 1998). This classification mainly suggested by Enzyme Commissions' Nomenclature Committee (IUBMB) (Tipton 1994; Barrett 1995; Barrett 1996; Barrett 1997) (**http://www.chem.qmul.ac.uk/iubmb/enzyme/EC34/**). Hence this classification has significance in a sense. Details of some of the general terminologies are given in figure 1.3.2.

I. Endopeptidase

Endopeptidases act on the alpha-peptide bonds situated away from the *N*-terminus or *C*-terminus. Some common examples are chymotrypsin, pepsin and papain. Endopeptidases have specific and limited role in proteolysis. Like, in removal of signal peptides from secreted proteins (e.g. signal peptidase I) and the maturation of precursor proteins (e.g. enteropeptidase) (Barrett and Rawlings 1991).

The endopeptidases are divided into sub-subclasses on the basis of catalytic mechanism, and specificity is used only to identify individual enzymes within the groups. These are the sub-subclasses of **serine endopeptidases** (EC 3.4.21), **cysteine endopeptidases** (EC 3.4.22), **aspartic endopeptidases** (EC 3.4.23), **metalloendopeptidases** (EC 3.4.24) and **threonine endopeptidases** (EC 3.4.25). Endopeptidases that could not be assigned to any of the sub-subclasses EC 3.4.21-25 were listed in sub-subclass **unassigned endopeptidases** EC 3.4.99 (Bergmann and Ross 1936; Rowan, Buttle et al. 1990; Barrett and Rawlings 1991; Rawlings and Barrett 1994; Rawlings and Barrett 1995; Rawlings and Barrett 1995).

II. Oligopeptidase:

Oligopeptidases tend to act on substrates smaller than proteins. Example of oligopeptidase is Thimet Oligopeptidase (Barrett, Brown et al. 1995; Knight, Dando et al. 1995).

III. Exopeptidase

The exopeptidases require a free *N*-terminal amino group, *C*-terminal carboxyl group or both, and hydrolyze a bond not more than three residues from the terminus (Hasegawa 1960; Nardi 1960). The exopeptidases are further divided into *N*-terminal acting peptidases, *C*-terminal acting peptidases and Dipeptidases. Each of them is defined as describe below.

i) *N*-terminal acting peptidases:

Aminopeptidases refers to EC 3.4.11 acts on the unblocked *N*-terminus of its substrate and release a single amino acid residue. Action site can be defined as: Xaa+peptide (or Xaa+Xaa$_n$). Examples are aminopeptidase-*N* and aminopeptidase-*C*. **Dipeptidyl-peptidase** refers in EC 3.4.14, hydrolyse an *N*-terminal dipeptide from its substrate. Action site can be defined as: dipeptide+peptide (i.e. Xaa$_2$+Xaa$_n$). Examples are dipeptidyl-peptidase I and dipeptidyl-peptidase III (Parsons and Pennington 1976). **Tripeptidyl-peptidase** also refer in EC 3.4.14, hydrolyses a tripeptide from the N-terminus of its substrate. Action site can be defined as: tripeptide+peptide (i.e. Xaa$_3$+Xaa$_n$). Examples are tripeptidyl-peptidase I and tripeptidyl-peptidase II (Doebber, Divor et al. 1978).

ii) *C*-terminal acting peptidases:

These peptidases hydrolyze a single residue, from the unblocked C-terminus of its substrate. Action site can be defined as: peptide+Xaa (or Xaa_n+Xaa). Examples are carboxypeptidase A1 and carboxypeptidase Y. Carboxypeptidases divided in sub-subclasses EC 3.4.16-18 in NC-IUBMB scheme. They are of **serine-type carboxypeptidases** (EC 3.4.16), the **metallocarboxypeptidases** (EC 3.4.17) and the **cysteine-type carboxypeptidases** (EC 3.4.18) (Rawlings and Barrett 1997). A **peptidyl-dipeptidase** refers to EC3 4.15 hydrolyze a dipeptide from the C-terminus of its substrate: peptide+dipeptide, and this explain the name. An example is peptidyl-dipeptidase A and angiotensin converting enzyme (Cushman and Cheung 1971; Lee, Larue et al. 1971).

iii) Dipeptidase:

Assign to EC 3. 4.13, hydrolyses a dipeptide, and typically requires that both termini are free: Xaa†Yaa. Examples are dipeptidase A and membrane dipeptidase.

IV. Omega-peptidase EC 3.4.19

This is the group of peptidases that have no requirement for a free N-terminus or C-terminus in the substrate. But their site of action remains close to one terminus or the other. Their action site is other than those of α-carboxyl to α-amino groups. Thus they are totally distinct from endo- or exo- peptidases. They can act on the terminal residues that are substituted, cyclized or linked by isopeptide bonds. Isopeptide bonds are peptide linkages other than those of a carboxyl to a -amino groups.

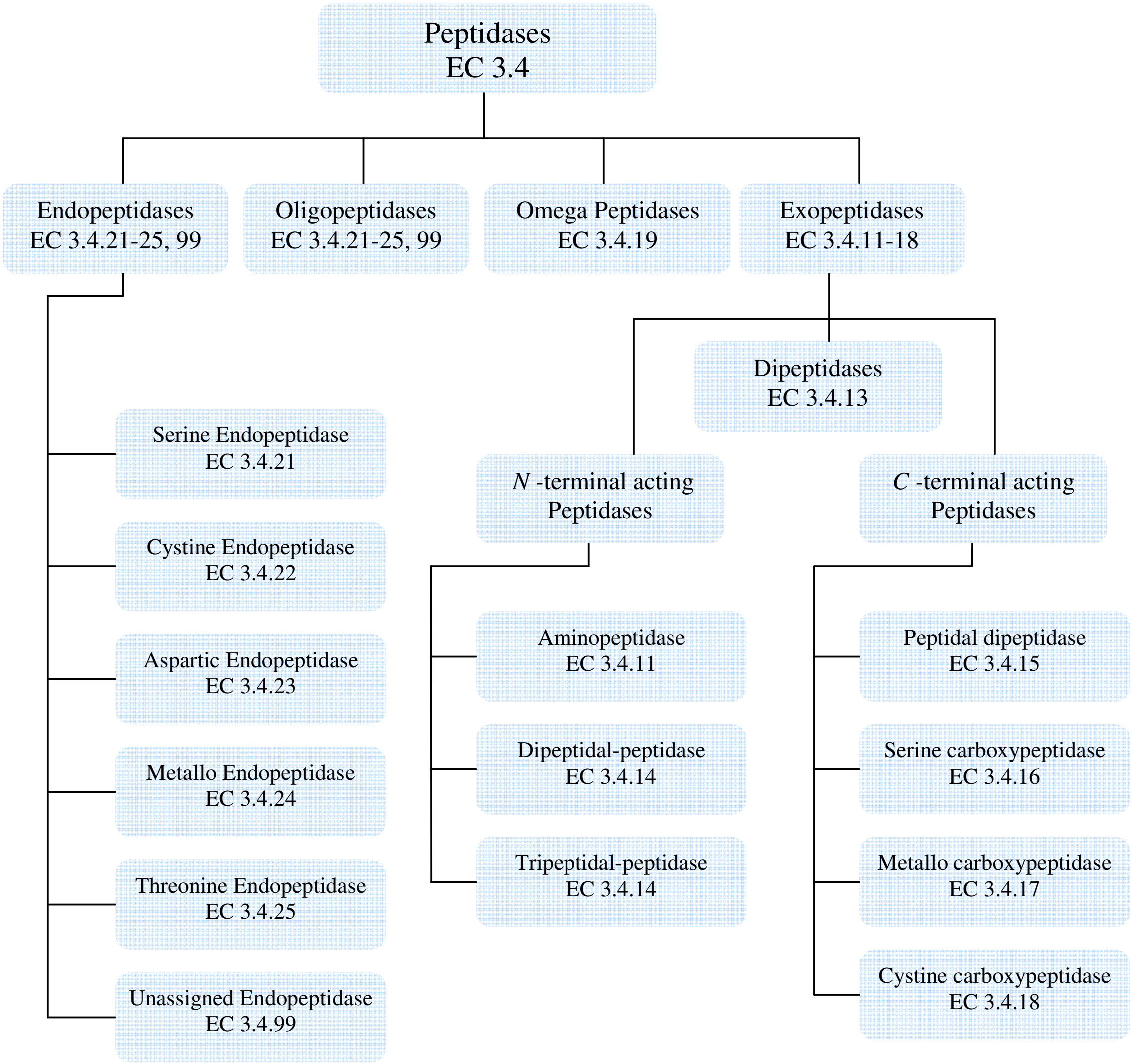


Figure 1.3.2: Enzyme nomenclature recommendations for peptidase from Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (**NC-IUBMB**).

1.3.3 Grouped by molecular structure and homology

This system was described in early 1990s (Barrett and Rawlings 1991). This is the most modern and organized system that correlates the individual entity from the evolutionary hierarchy and categorizes them in similar protein sequence "families". In 1993, Rawlings & Barrett described a system in which individual peptidases were assigned to species of peptidase (Sub-Family) and homologous peptidases from different peptidal species merge to made families, and the families were further grouped in clans according to their 3D structure homology (Rawlings and Barrett 1993). This scheme was developed to provide the structure of the MEROPS database (figure: 1.3.3) and extended to include the proteins that inhibit peptidases (Rawlings and Barrett 1999; Rawlings and Barrett 2000; Barrett, Rawlings et al. 2001; Rawlings, O'Brien et al. 2002; Rawlings, Tolle et al. 2004).

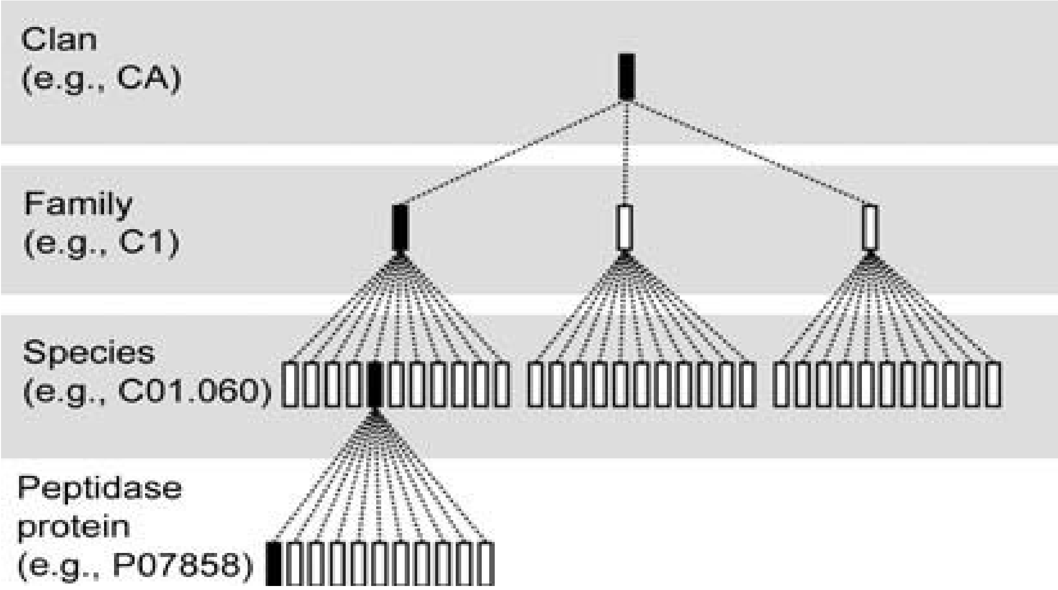


Figure 1.3.3: schematic view of MEROPS database system (Barrett and Rawlings 2007).

This classification system was unique because it has taken all of the protein in consideration that contains the peptide domains as their functional unit, it excludes homologous domains with non proteolytic functional units (Barrett, Rawlings et al. 2001).

I. Family: peptidase family is group of peptidase activity residing proteins that are closely homologous in sense of amino acid sequence. Each family has based around the founder member of family called “type example”. The significant homology of whole proteins a.a. sequences or domain of any protein that has essential catalytic resides can be included to the “type example group” called Family (Barrett, Rawlings et al. 2001). The homology to ‘type example’ was detected by blast against non redundant databases (Altschul, Gish et al. 1990). The criteria of strict homology was chosen as '*e-score*' less then 0.01, as describe by Reek et al. (Reeck, de Haen et al. 1987).

Each family is identified by an alphabetic capital letter representing the catalytic type followed by unique number assigned to family.
Like **A** for aspartic-type, **C** for cysteine-type, **G** for Glutamic-type, **M** for metallo-type, **S** denotes serine-type, **T** for threonine-type; **U** for unknown-type and **X** for compound-type catalytic peptidases.

II. Clan: Evolutionary relationship or evidence of common ancestry revels by three dimensional structure of proteins. Since tertiary structure of every protein is not known and not easy to compare, secondary structure elements of the 'type example' are used to compare order of catalytic-site residues in the polypeptide chain and by common sequence motifs around the catalytic residues (Rawlings, Morton et al. 2006). Secondary structures calculated from the PDB file according to Kabsch and Sander and converted to GIF image by Perl Script (Kabsch and Sander 1983). This GIF image provide sufficient over-look to relate families into clans in absence of crystallographic data (Barrett, Rawlings et al. 2001).

Each clan is identified by two alphabetic capital letters, the first specifying the catalytic type and the second are unique to the clan. The first letters used in clan are **A-**Aspartic, **C-**Cysteine, **G-**Glutamic, **M**-Metallo, **S**-Serine, **T-**Threonine, **U-**Unknown and **P-**Mixed or compound catalytic type of peptidase clan.

Table 1.3.3: Statistic from http://merops.sanger.ac.uk/ (*MEROPS Release 8.00), update 4th Aug2008)*

TOTALS FOR ALL CATALYTIC TYPES				
Catalytic Type	**Sequences**	**Identifiers**	**Identifiers with EC numbers**	**Identifiers with PDB entries**
Aspartic	4667	199	31	38
Cysteine	1687	633	54	101
Glutamic	41	5	2	1
Metallo	35579	694	127	101
Serine	36929	996	112	165
Threonine	3727	76	21	24
Unknown	2994	24		1
Grand Total	100807	2627	347	431
Total families				202
Total clans				51

1.4.0. Protease Specificity subsite:

Berger has introduced a model system for describing the specificity of peptidases. According to this model a catalytic site is considered to be flanked on one or both sides by specificity subsite, each able to accommodate the side chain of a single amino acid residue (Abramowitz, Schechter et al. 1967; Schechter and Berger 1967; Berger 1970). These sites are numbered from the catalytic site of enzyme as, S1...S*n* towards the N-terminus of the substrate, and S1'...S*n'* towards the C-terminus. The residues they accommodate from substrate are numbered P1...P*n*, and P1'...P*n'*, respectively. In this representation, the catalytic site of the enzyme is marked *. The peptide bond cleaved (the scissile bond) is indicated by the symbol '†' or a hyphen in the structural formula of the substrate, or a hyphen in the name of the enzyme (figure: 1.4.0.A).

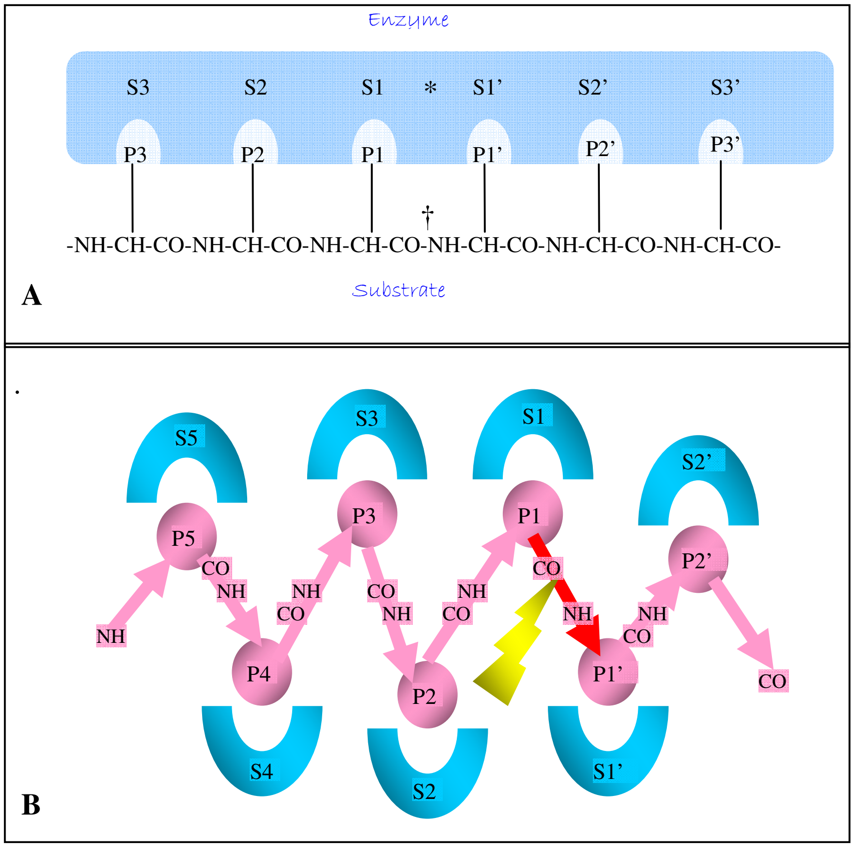


Figure 1.4.0: (A) P' denotes the a.a residues from substrate and S' denotes interacting a.a residues from Enzymes, † denotes cleavage site. (B) 2D representation of protease specificity subsite. Derived from (Timmer and Salvesen 2007)

In this respect, it is important to know that pattern of residues from substrate are represented from a single continued polypeptide chain, while residues from enzyme may come from the different fold of protein complex, but they should have to be arranged in an appropriate three dimensional spatial position to allow only the specific substrate residue confirmation to enter into the specificity pocket and interact with active site residues during the cleavage action (figure: 1.4.0.B).

Example of such substrate binding in specificity pocket and then consequent availability to active site residues interaction can be describe by the trypsin like protease family. Between the two domains there is an active catalytic triad that right on top of the specificity pocket, made up of two different loops that determine primary specificity. While, other loops near the specificity pocket determine secondary specificity (inability to accept certain residues or preference to accept specific residues).

Table 1.4.0: Specificity of some selected proteases

Enzyme	P2	P1	P1’	P2’	References
Chymotrypsin low specificity		F/L/Y	not P		(Gasteiger, Hoogland et al. 2005)
		W	not M/P		
		M	not P/Y		
		H	not M/P/ D/W		
Chymotrypsin high specificity		F/ Y	not P		
Trypsin affinities		K/ R	not P		(Keil 1992)
	W	K	P		
Trypsin Exceptions	M	R	P		
	C/ D	K	D	-	
		C	K	H or Y	
		C	R	K	
		R	R	H or R	
Elastase		A/L/I /G/S/V			(Grunnet and Knudsen 1983; Bode, Meyer et al. 1989)

1.5.0. Commercialization potential of Proteases:

In 2004 total annual world market for enzyme was US billion 1.5 dollar and is estimated to increase at the rate of 6.9% per annum. Proteases alone generate half of the total sale revenue generated by all industrial enzymes (McGrath 2005). But the countries like USA where Corn is utilizing as biofuel, protease demand is next to carbohydrates and it considers as higher value enzyme with greater net gross profit industrially applicable product. (Hayes, Zimmerman et al. 2006).

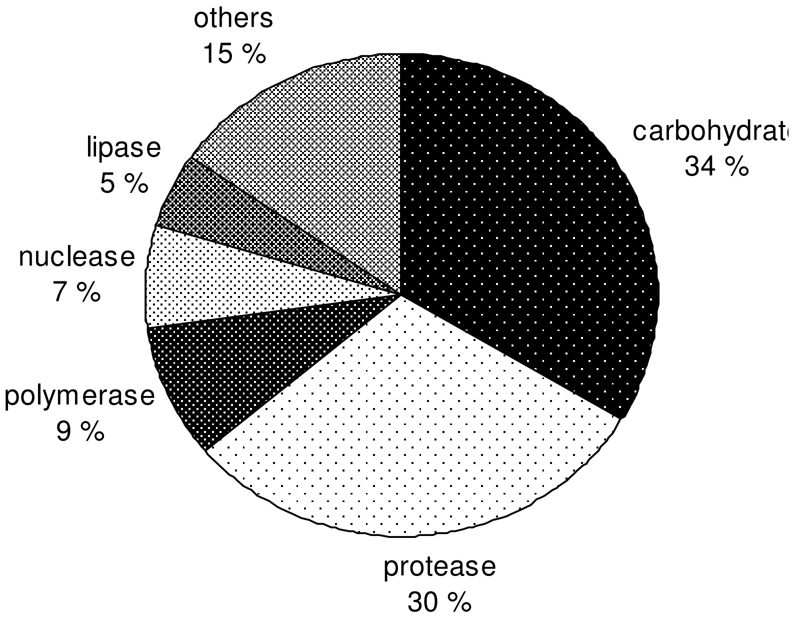


Figure 1.5.0: US Enzyme demands, expected for the year 2015. (Source: the freedonia Group)

According to the US Enzyme market survey 1995 to 2015 (Hayes, Zimmerman et al. 2006) protease demand in terms of its worth, is covering more than half of the all enzymes market in USA. Cleaning and laundry purpose proteases which are much in numbers are also estimated to be high income generating industries. Food and beverages processing protease enzymes demand comes right after it, especially in dairy industries.

1.5.1. Pharmaceutical applications:

One of the major examples of the pharmaceutically applicable protease enzyme includes thrombolytic enzymes. Blood clot or fibrin formed abnormally, in case of some life threatening disease like **AMI** (acute myocardial infarction), **PAO**

(Peripheral arterial occulusion), **pulmonary embolism**, and ocular catheter formation. This phenomenon is balanced in the body by plasmin (EC 3.4.21.5), which is activate from plasminogen with the help of tissue plasminogen activator (t- PA).

In case of imbalance in homeostasis, fibrin can not be lysed and cause strokes in heart, while in brain causes break in oxygen supply resulting in cerebral infarction and dementia (Sugimoto, Fujii et al. 2007). Numbers of enzymes are known in this series which have been used and modified for example Streptokinase, Urokinase from first generation that have been now modified to Anistreplas, Altaplase, Tenectaplase, Reteplase etc. from second generation with more specificity and accuracy (Hayes, Zimmerman et al. 2006).

Another important example in the area of pharmaceutically applicable proteases includes Botulinum toxins A and B that are known to have a potentially neuromodulating effect when applied in controlled manner. It acts on the synapses of nerve cells and cleaves key protein needed to transmit nurotansmitter acetylcholine across nerve cell membrane, resulting in localized paralysis. This characteristic can be exploited in epilepsy, uncontrolled muscles motions, prostatic hyperplasia for cosmetic purposes like reducing wrinkles, to simply give relax to muscles, stopping the sweat under arms and on palms and for the treatments of Blepharospasm (involuntary or spasmodic twitching of certain eye muscles), Strabismus (inability of focusing both eyes on same objective).

Its demand is progressing with the rate of 13% per annum and expected to reach 220 million USD in 2010. Proteases have also a great prospect in the area of digestive enzymes used for therapeutic purposes. They are used for malabsorption for people who suffer from digestive problems (Hayes, Zimmerman et al. 2006; Sievert, Bremer et al. 2007; Wefer, Seif et al. 2007; Boy, Seif et al. 2008; Seif, Boy et al. 2008).

1.5.2. Digestive Enzymes:

Most of the protease enzymes fall in the category of digestive enzymes when considering the Enzyme Replacement Therapy (HRT). Patients suffering from the deficiency of digestive enzymes not only have to suffer from malnutrition but also suffer from the other life hazarding diseases like cystic fibrosis, chronic pancreatitis and pancreatic cancer. Protease drinks have been introduced to market that provides a combination of protease mineral and vitamins and claims to relief in inflammation for the addition of Bromelain and other known orally administrable proteases of pleasant health effects.

1.5.3. Food & Beverage applications:

Proteases are also been in use in the food and beverage industry where certain flavors and texture are assisted with the use of proteases. For example in bakery it is combined with the transglutaminase, and with beta glucanase in winery for better texture and flavor. In food industry the major role of these enzymes have in meat and fish processing where they act as meat tenderizer and flavoring agent some time by forming the glutamic acid and hence replace the monosodium glutamate. Among other implications, soybean milk processing and wheat gluten processing industries are also major application areas.

For the cleaning purpose much of the proteases share belongs to the laundry applications where research is in full support to find out extremophilic nature protease that can act on higher/ lower temperature, higher/ lower pH and can be oxidativly stable to be able to work in the presence of bleach.

Protease applications in contact lens solutions and facial masks, skin cleaners, hair removing creams and faster teeth cleaning past and mouth cleaning hygiener are well known. Papain, bromeline, trypsins, pancreatin and collagenase with their exfoliating properties, holds the potential cosmetic applications.

In leather industries, protease assistance is used to dissolve certain proteins to remove scud, promotes the opening up of skin substance to enhance dying properties, to make the leather pliable and increasingly soft and cleaner. Proteases also add up well with the catalyst in de-hair process of leather with minimal damage to grain.
Summary of these marketable areas are listed in table 1.5.3.

Table 1.5.3: United States Protease marketing analysis and prediction from year 1995 to 2015. Data presented in US Million Dollars (Hayes, Zimmerman et al. 2006).

	Market Area	YEAR				
		1995	2000	2005	2010	2015
I	Pharmaceutical Market for protease	118	139	235	386	596
a)	Digestive enzyme market for protease	21	26	29	38	79
b)	Cardiovascular Thrombolytic Protease Market	80	64	65	93	118
II	Food & Beverages Processing protease market	48	58	70	81	93
a)	Dairy processing Protease market	27	32	37	41	45
i)	Chymosin/ rennet market (do not match with the data given in 2nd Table)	25	29.4	34.3	37.2	40.7
III	Cleaning purpose Protease market	84	95	94	107	122
a)	Laundry applicable Protease market	81	89	85	95	106
b)	Dishwasher applicable Protease market	3	6	9	12	16
IV	Research Biotech applicable Protease market	8	12	14	18	23
V	Cosmetics & Toiletry applicable Protease Market	7	11	15	19	23
VI	Textile & leather processing protease market	15	13	9	9	9
	Over all Protease Enzyme demand	292	345	460	653	915
	Total other Enzymes demand	930	1190	1605	2240	3020

1.6.0. Sources of proteases:

Proteases are commonly been purified mostly from plant source such as papain from papaya and bromelain from pineapple. Rennet, pancreatic proteases have been isolated from animal sources, but now with the introduction of cloning their use in the commercial companies have been declined. Microbial sources have been investigated form several years in the past and many novel proteases have been discovered from the pathogenic and extremophile sources.

With respect to the significance of proteases as profitable industry products, many biotechnological industries have focused their concerns to produce them through DNA recombinant technology. Special focus is with the kind of protease which can remain active on broad pH range and have the temperature stability at unusual temperature. Cloning or DNA recombinant technology by choosing microorganisms as the host organism is the most desirable to produce enzyme for commercial purpose because:

- In this technology genetic alteration or direct evolution can be used systematically and predictably to enhance the function of enzyme for certain applicable area (Carter, Dunn et al. 2008; Di Cera 2008).
- Recombinant technology is also desirable to be able to overcome the rareness of source organism.
- Proteins from extremophiles can be managed to be produced in laboratory conditions.
- In recombinant technology microbial cultures are faster to grow compared to plant or animal source, so after a certain initial time of recombinant production it is much time saving to use recombinant microbes for every batch harvest.
- Reproducibility with no genetic variation in a controlled system is another important advantage when compared to natural source utilization.

According to the survey report of Fredonia Group microbial recourses are major source which are expected to provide the 61.3% of total protease demand in year 2015.

Table1.6.0: United States' Type of Protease demand analysis and prediction from year 1995 to 2015. Data presented in US Million Dollars and represented in percentage as well (Hayes, Zimmerman et al. 2006).

	Type of Protease	YEAR				
		1995	2000	2005	2010	2015
	Total protease demand	292	345	460	653	915
	% of total enzyme demand	31.4%	29%	28.7%	29.2%	30.3%
I	Microbial protease demand	125	167	257	385	561
	% of Total Protease demand	42.8%	48.4%	55.9%	59%	61.3%
II	Fibrinolytic Protease demand	84	70	72	102	130
	% of Total Protease demand	28.8%	20.3%	15.7%	15.6%	14.2%
III	Chymosin/ rennet demand	36	43	51	58	66
	% of Total Protease demand	12.3%	12.5%	11.1%	8.9%	7.2%
IV	Other Protease demand	47	65	80	108	158
	% of Total Protease demand	16.1%	18.8%	17.4%	16.5%	17.3%

1.7. Protease Biological functions:

Proteases are important through out the life from conception, birth, life, ageing, to death of all organisms (Abraham and Potter 1989; Lala and Graham 1990). They are biology's version of Swiss army knives, cutting long sequences of amino acids (called peptides) into fragments that fold into functional proteins (Seife 1997). Proteases are interesting molecules that posses harmful and beneficial characters at the same time, they posses catabolic and anabolic characters, they can cause diseases and sever damage to body if inserted from other organisms on one hand but on the other hand some protease from other organisms are used to give a disease relief and healthy life. A slight change of their balance in body (hypo or hyper activity) can cause disease symptoms. There are almost 50 known human genome sequences in which single mutation can cause the genetically or hereditary disease resulting in the over/ under production of protease or protease inhibitor/ activators, leading to pathological condition (Puente, Sanchez et al. 2003). Proteases are the main focus of modern research to cure several diseases; some of the examples are given in table 1.7.0.

Table 1.7.0: Protease related diseases in Human beings.

Diseases	Protease involved	Reference
Human Immuno Defficiency Virus infection	HIV-protease	(Goldberg and Stricker 1996; Perryman, Lin et al. 2006)
Blood cancer	T cell leukemia virus	(Li, Laco et al. 2005)
Alzheimer's disease	amyloid-beta peptides	(Beher and Graham 2005)
hemostasis, repair, cell survival, inflammation, and pain Protease activating G-Coupled protease	G- coupled Protease-Activated Receptor (PAR)	(Kawabata 2001; Ossovskaya and Bunnett 2004)
Kawasaki disease	PMN-derived elastase,	(Saji 2008)
Cardiovascular diseases	Serine protease Corin	(Wu 2007)
Foot and mouth diseases	3C virus protease	(Curry, Roque-Rosell et al. 2007)
Osteoarthritis diseases	proteases	(De Nanteuil, Portevin et al. 2001)
Aging, hereditary cerebral hemorrhage, Alzheimer’s down syndromes	Alpha 1-antichmotrypsin	(Abraham and Potter 1989)
Air ways injuries	proteases	(Rennard, Rickard et al. 1991)
Asthma, allergies	Bromelain, papain	(Baur and Fruhmann 1979)
Cancer, breast cancer, apoptosis irregulations, cancer metastasis	proteases	(Hocman 1992; Kennedy 1993; Das and Mukhopadhyay 1994; DeClerck and Imren 1994) (Rochefort, Capony et al. 1990; Rochefort and Liaudet-Coopman 1999)

According to a careful estimation of Southan, C. protease comprises ~1.8% of the human genome, and genome data annotation revels protease inhibitor ratio as 10:1 (Copyright © 2000 European Peptide Society and John Wiley & Sons, Ltd.). This estimation ended with 700- 1100 proteases and 70–110 protease inhibitors. Protease comprises little higher almost 5% of genomes of infectious organisms (Southan 2000).

In most of the pathogenic organisms, proteases are the major causes of pathogenesis (Travis, Potempa et al. 1995). It severity range from mild fever, or pain to even death, like in case of *cane disease* caused by *Clostridium botulinum*. When pathogens invade there is a sensation of pain in the result of dysregulation of kallikrein and kinin pathway (figure: 1.7.0.) caused by bacterial protease (Nilsson, Carlsson et al. 1985; Maeda and Molla 1989).

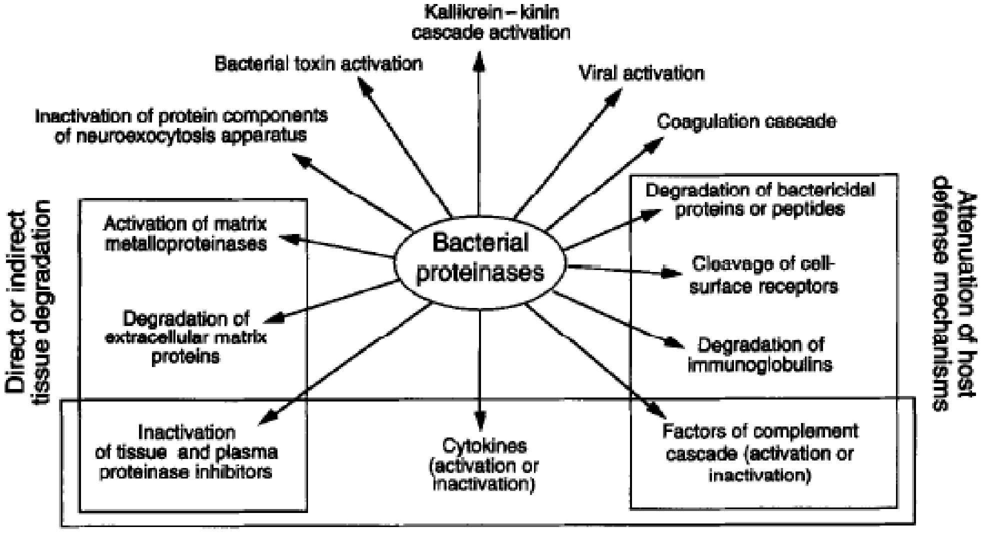

Kallikrein – kinin cascade activation
Bacterial toxin activation
Viral activation
Inactivation of protein components of neuroexocytosis apparatus
Coagulation cascade
Direct or indirect tissue degradation
Attenuation of host defense mechanisms
Activation of matrix metalloproteinases
Bacterial proteinases
Degradation of bactericidal proteins or peptides
Cleavage of cell-surface receptors
Degradation of extracellular matrix proteins
Degradation of immunoglobulins
Inactivation of tissue and plasma proteinase inhibitors
Cytokines (activation or inactivation)
Factors of complement cascade (activation or inactivation)

Figure 1.7.0: Functions of bacterial proteinase in infection
Source: (Miyagawa, Nishino et al. 1991)

They mainly act as degradative enzymes among the lower to higher organisms and in some instance they act very specifically and help the cell in the vital function of protein folding, cell signaling, hormonal communication etc. therefore deficiencies of theses enzymes in biological system can convert into diseases. Concerning these biological effects of protease deficiencies, **enzyme therapy** for protease is in practice. Numbers of orally administrable protease drinks prepared from natural sources are available at the market. Several reasons have been describing in this respect (http://www.enzymeessentials.com/HTML/protease.html; http://www.enzymeresearchgroup.net/protocols.php). It has been described that proteases, when taken orally can be absorbed by alpha2 macroglobulins. They used to encounters dead, damage and foreign unidentified protein particles of allergens and pathogenic factors from bacteria, fungi, insects and other organisms.

1.8.0. Introduction to the organism – *Aliivibrio salmonicida*:

A. salmonicida LFI1238 is a halophilic ("salt loving") and psychrophilic ("cold loving"), curved, gram-negative bacterium (Hoff 1989; Colquhoun and Sorum 2001). It is known as the causative agent of "cold-water vibriosis (CV)" or "Hitra disease", occuring at low temperature and causes hemolysis and tissue degradation in fishes (Salte, Nafstad et al. 1987; O'Halloran 1993; Stephen 1993). It was predominant in winter time with low water-temperatures (Holm and Jørgensen 1987). In contrast to other septicaemic, hemorrhagic, pathogenic bacteria, no exotoxin have been identified until now (Hjeltnes, Andersen et al. 1987; Holm and Jørgensen 1987). For this reason *A. salmonicida* is an interesting model organism for the study of temperature and host adaptation mechanism.

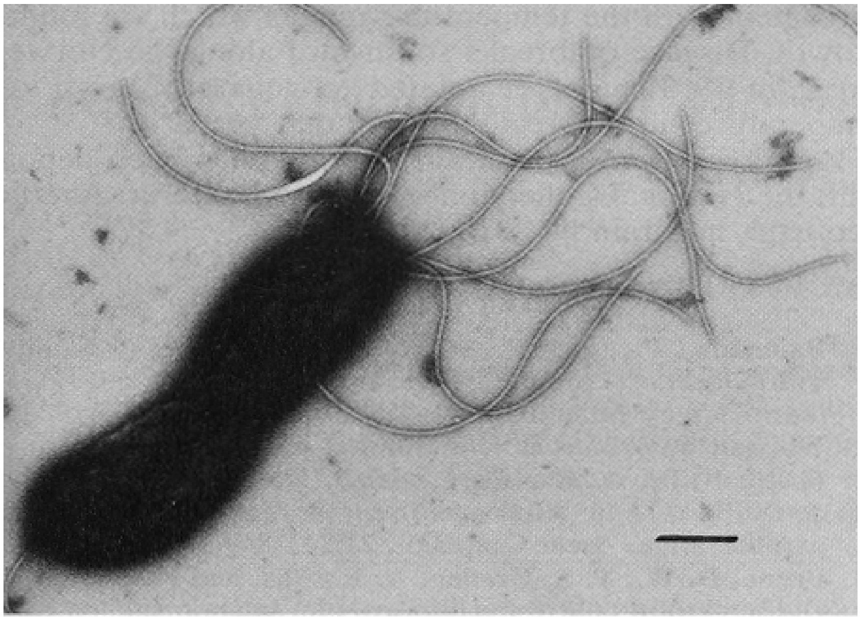


Figure 1.8.A: Electron Microscopic photograph of *Aliivibrio salmonicida*
(Photo taken by Steinar Paulsen, Protein Research Group, UITø, Norway)

Psychrophiles have the ability to survive and proliferate at low temperatures. They have been modified under the constant cold environments challenges. These organisms and their building blocks (proteins) posses the quality that distinguish them from the organism that can not survive in cold environment. D'Amico et al. described some of these challenges like reduced enzyme activity, decreased membrane fluidity, altered transport of nutrients and waste products, decreased rates of transcription, translation and cell division, protein cold-denaturation, inappropriate protein folding, and intracellular ice formation. Cold-adapted organisms have successfully evolved features, genotypic and/or phenotypic, to surmount the negative effects of low temperatures and to enable growth in these extreme environments (D'Amico, Collins et al. 2006).

To study the genome of *A. salmonicida* a project was organized, by department of molecular biotechnology, UiTø and NorStruct. The genome consists of two chromosomes, two megaplasmids and four plasmids. Shot gun libraries have been constructed in collaboration with, the Welcome Trust Sanger Institute, to sequence the whole genome, (www.sanger.ac.uk/Projects/V_salmonicida/). This genome have been recently sequenced and published (Hjerde, Lorentzen et al. 2008). (Figure: 1.8.B)

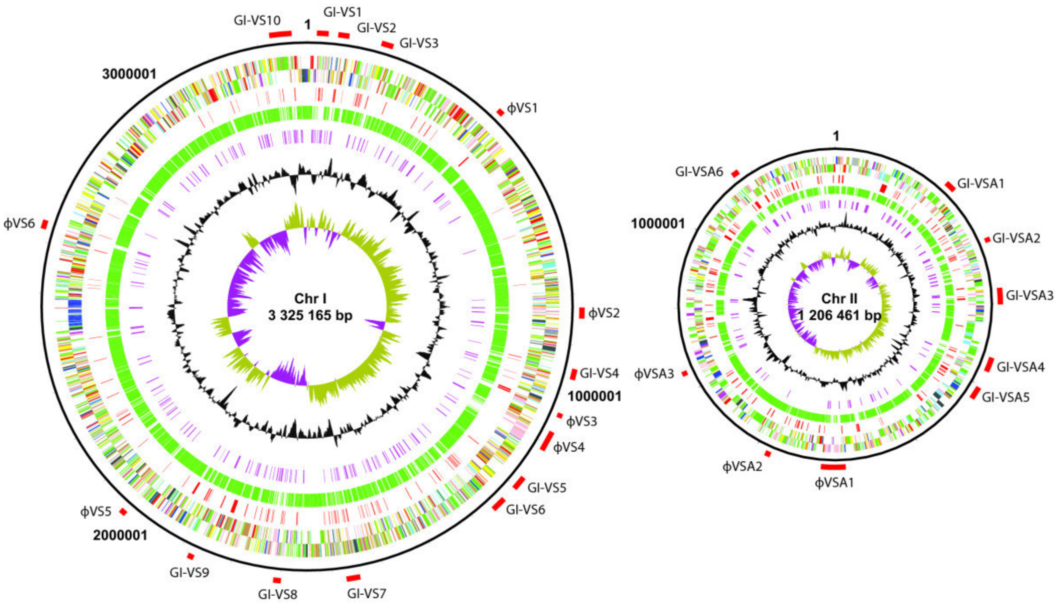


Figure 1.8.B: chromosomal circular diagrams (outside to inside)): scale (in Mb), unique CDSs compared to the other *Vibrionaceae* species (red), orthologues shared with the other *Vibrionaceae* species (green), IS element transposases (purple), dark blue, pathogenicity/adaptation; black, energy metabolism; red, information transfer; dark green, surface associated; cyan, degradation of large molecules; magenta, degradation of small molecules; yellow, central/intermediary metabolism; pale green, unknown; pale blue, regulators; orange, conserved hypothetical; brown, pseudogenes; pink, phage + IS elements; grey, miscellaneous. The positions of phage elements and GIs larger than 5 kb are marked (red); source: (Hjerde, Lorentzen et al. 2008)

1.9.0. Protein expression through cloning:

Proteins in host organisms express in a limited quantity and usually in the response of certain stimuli. The main objective of the cloning is to obtain the elevated level of target protein expression, higher to the source organism of target protein, so that it can be obtains in milligram quantities necessary for structural and functional characterization. Cloning systems can be utilized in a controllable manner to expresses the target protein in response to the stimulants, when ever needed. These systems require the inducer or stress condition to unblock the progressive transcription of cloned gene of target protein.

With the development of cloning and protein expression technology, numbers of choices are present not only for cloning methodology but also for cloning systems that composed of expression vector and expression clone (choice of cell types for expression). Careful selection of cloning system according to the protein characteristics and tag requirement for protein purification is the major and initial step in protein expression through cloning technology. The second scaling up step is the optimization of growth conditions or media constituents for highest possible soluble yield in a single batch. As a whole, theoretical and experimental decision in these two steps govern the economically feasible expression system development.

1.9.1. Bioinformatic analysis of protein:

Before trying to express a target protein a general bioinformatics analysis of protein is required to develop a strategy for purification for example PI calculation in case of ion exchange chromatography. A model building is also helpful to guess the exposed terminal for tag attachment (N/C-terminal tag) for purification purpose. Closely related proteins information is also useful to design a cloning/ purification strategy and possibility of heteromer formation. Some proteins are designed in nature for extra-cellular or periplasmic expression, these protein can be detected by analysis through SignalP (Bendtsen, Kiemer et al. 2005; Emanuelsson, Brunak et al. 2007).

Such proteins usually express when transported out of the cell or in periplasm. Such target proteins usually exported out with the help of the signal sequence from the expression vector and cloned without signal sequence. Some proteins are also expressed in zymogene form, in that case pro sequence needed to be identified by aligning with an enzyme of same category, when needed to be express in the active form. These proteins by understand the Half life, THMMH, solubility

1.9.2. Primer designing:

Primers should be design in such a way that it will bring the gene code of target protein in frame with the tag and initiation codon. In spite of all the precautions possibility of missing frame or any mutation during the purification from the gel (by excess UV exposure) can be indicated with the sequencing of cloned gene.

1.9.3. Selection of cloning technology:

With recent advancement in Cloning science, traditional **cDNA** cloning method is now replacing several types of high throughput cloning methods. Numbers of factors are important in their selection that includes, high fidelity (assurance), ease of use, reliability of system, validation of correctly cloned system, flexibility to change the express species or vector and overall cost of recombination, time consumption (Marsischky and LaBaer 2004).

I. Traditional cloning:

Traditional cloning methods involve restriction, ligation enzymes, planed cautiously for individual clone; therefore it lacks the universality. Absence of flexibility to switch over between vectors and expression cells make it costly in sense of time consumption hence proved undesirable for thigh throughput structural studies.

II. Gateway cloning:

High throughput cloning method based on site-specific recombination properties of bacteriophage Lambda. Since this method is efficient in switching between different expression clone therefore it is most desirable in expression studies (Walhout, Temple et al. 2000). The presence of ccdB gene (killer gene) in donor and destination vectors gives >95% assurance of recombination process. This system requires the expression host that carries the lysogenic T7 RNA polymerase gene in their genome that expresses upon induction and cross activate the transcription of inserted gene in expression clone (figure 1.9.3)

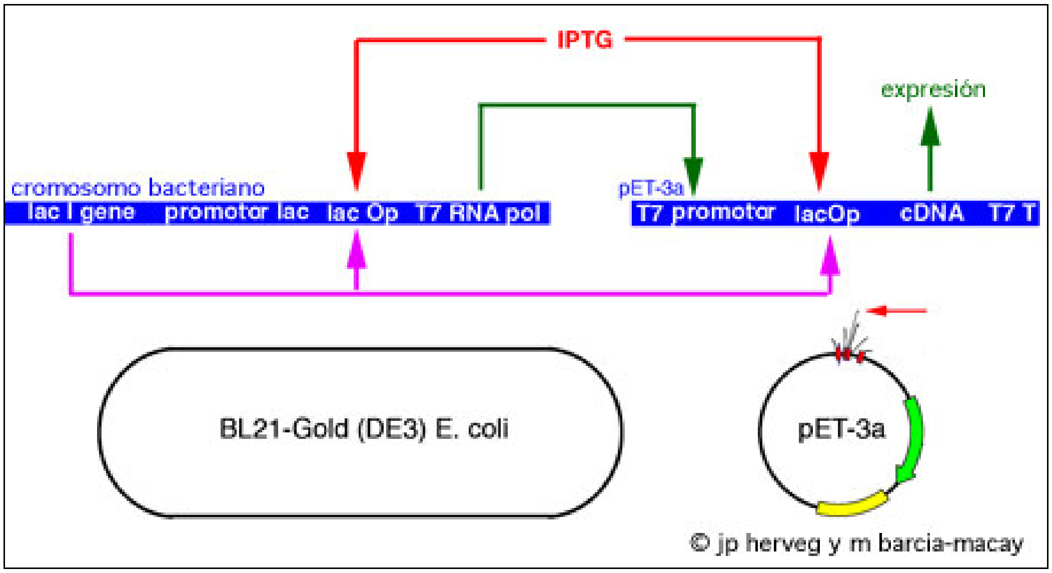


Figure: 1.9.3: mechanism of expression in Gateway system™.

This system bears all desirable qualities including high fidelity, reliability, flexibility and universality, therefore most prevalent among molecular biologist and biochemist.

III. In-Fusion/ Ligation independent cloning (LIC):

LIC is an efficient and intelligent high throughput cloning system independent to the use of bacteriophage λ specific attachment (*att*) site. It utilized 12-nt cohesive ends from amplified PCR product and vector with compliment sticky end. Hence ligation occur independent of ligase enzyme in a brief *in vitro* incubation. Selection of cloned identified as white colonies from a IPTG and X-Gal provided plates, since insertion of gene disrupt the lacZ gene (Aslanidis and de Jong 1990; de las Rivas, Curiel et al. 2007; Tachibana, Tohiguchi et al. 2009).

IV. Precision cloning:

This method is a recent advancement of recombination method, where it implies to reduce the linker sequences that often remains attached with cloned gene, and afterward in expressed protein. Hence this system is modified to avoid the drawbacks of the two above mentioned systems (Engler, Kandzia et al. 2008).

1.9.4. Selection of expression vector:

The role of expression vector is to provide suitable tag/tags to facilitate the solubility and fast recovery or capturing in the isolation procedure. These expression vectors are provided with the antibiotic resistance genes for the selection of plasmid containing cells after the cloning and during the expression trials. There are almost hundreds of choices for selection of expression vector depending on their variable traits like, origin of replication, copy number, promoter systems, induction methods, compatibility with host strains, control over expression, yielding capability, choices of fusion partner etc.

Decisions made at this stage have significant effect in overall performance of the protein expression project, in term of overall cost and efficiency of the system. The traits concerning the factors described above will be discussed further in the following sections.

1.9.5. Choice of promoters (gene regulation):

Expression vectors are provided with the main governing machinery for controlled operation of protein expression. This machinery is called "inducible promoter", which is usually well studied for expression control in response of certain physical or chemical stimuli. Operons are the cluster of functionally related structural genes under the regulations of regulatory genes (operator sequence) situated within a promoter sequence. (Jacob, Perrin et al. 1960; Jacob and Monod 1961). Transcription of these structural genes in a polycistronic mRNA (single RNA coding for several structural genes) can be switch ON or OFF, in response to certain stimuli, depending on the cells metabolic needs.

The operators are of two types, **Repressible Operon** and **Inducible Operon.** Repressible operons usually belong to anabolic processes, they remain turned on until the repressor is activated with the biosynthesis end product, such as ***Trp Operon*** (which belongs to the regulation of structural genes for the synthesis of tryptophan). Inducible Operon usually belongs to a catabolic processes, they remains turned off until a pathway main substance deactivates the repressor, such as ***Lac Operon*, *Ara***

Operon, which regulates the structural genes for catabolism of Lactose and arabinose respectively (Campbell and Reece 2006).

There are many expression vector known, emerging from the modification of the main roots of few well studied operating systems like Arabinose Operating system (*P_{BAD} / P_{ARA} Operon*), Lactose Operating System (*Lac Operon*), Tryptophan operating system (*tac Operon*), phage operating systems (T_7, T_5) etc. Numbers of newly emerging promoter systems are also coming in market as research is progressing toward the discovery of unique gene regulating systems, the table 1.9.5 summarize them with their advantages and drawbacks (Cabrita and Bottomley 2004; Jana and Deb 2005).

Table 1.9.5: Bacterial promoters use in expression vectors.

Promoter	Regulation	Induction	Drawbacks	advantages	References
lac (lac UV5)	Lac I, lac I^q	IPTG, thermal	Leaky expression, Low level expression compare to other system	Wide verities of vectors available	pTrip1Ex2 (clontech)
T7	Lac I, lac I^q	IPTG, thermal	Leaky expression, difficult to achieve high cell density	multiple tags avalible, specific T7 RNA polymerase expression	pET (Novagen)
T5	Double lac promoter	IPTG	Limited vectors(12) Limited strains (2)	Very tight regulation, native *E.coli* RNA polymerase expression	pQE, (Qiagen)
tac	(trap + lac operator) Lac I regulation	Thermal > IPTG	Leaky expression, fewer fusion tag available	Huge expression at higher temperature, use *E. coli* RNA polymerase	pMAL (New England Biolabs), PGEX (
trc	-	IPTG	-	-	pTrcHis (invitrogen)
trp	-	Trp, starvation IAA	Leaky expression	-	pLEX (invitrogen)
Prha	rhaBAD	rhamnose	-	-	(Wilms, Hauck et al. 2001)
Ara	araC	L-arabinose	Few vectors available, expression repressed with glucose	Tight regulation, expression of toxic proteins	pBAD (invitrogen)
$Pzt1/P_{LtetO-1}$	Tet R/O	tetracycline	-	-	pLP-PROTet-6xHN (clontech)
λpL	λcIts 857	Heat shock 42°C	Expression cannot be brought at low temperature	Huge expression for thermal adapted proteins	pKC30 (invitrogen)
RpoS	σ(suS)	Cold shock 15-20°C	Repression is not fully controlled	Good for cold adapter proteins	(Baneyx and Mujacic 2002)
phoA	phoB, phoR	Phosphate starvation	Limited media options, not titratable		pBKIGF2B-A
Cad	cadA	Low pH	Limited characterization, less vectors available	40-50% of total biomass, Good for acidophilic proteins	pSM10, (Chou, Aristidou et al. 1995)
RecA	lexA	Nalidixic acid	Not titratable.	Good for acidophilic proteins	-

In addition to above described bacterial promoter systems, yeast, insects and mammalian cells have their own different type of promoter for controlled expression. *GAL1, Gal4 AD/BD* (invitrogen™) are promoter systems for yeast cells efficient expression. In insect cells, *Polyhedrin* promoter is used for baculovirus based expression, while *DES MT* and *OpIE2* promoters are used in non viral expression (invitrogen™). In mammalian cells based expression CMV promoter involves with transduction based expression, EF-1a involves non viral mammalian cell expression (invitrogen™).

I. Lac Operon:

Lac operons are well studied operon systems that are used in enteric bacteria for lactose catabolism when glucose is not available as energy source. They are very common in use for the cloned genes' controlled regulation and efficient transcription (Simons, Houman et al. 1987).

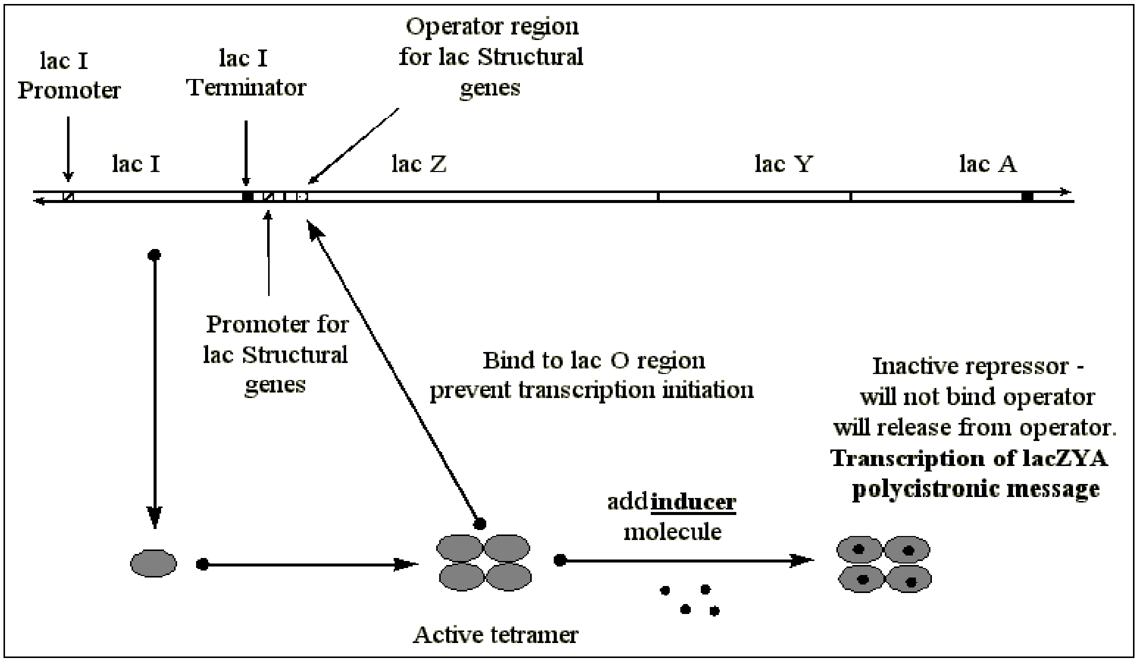

Operator region for lac Structural genes
lac I Promoter
lac I Terminator
lac I
lac Z
lac Y
lac A
Promoter for lac Structural genes
Bind to lac O region prevent transcription initiation
Inactive repressor - will not bind operator will release from operator.
Transcription of lacZYA polycistronic message
add inducer molecule
Active tetramer

Figure1.9.5.A: depiction of Lac Operon controlling system. Source: (Blaber 1998)

In this system LacI is the repressor of the Lac structural genes promoter region. It synthesizes constitutively from upstream of the LacZYA gene/inserted gene (incase of expression plasmids), and binds from the promoter region, hence block the RNA polymerase active transcription. When the lactose is provided to the media, an isomer of lactose, Allolactose or analogs (IPTG) can bind specifically with the repressor and decreases its specificity from the promoter region. By this the active transcription starts (figure 1.9.5.A). IPTG is most commonly use in the cloning system as inducer of Lac Operon, since it can efficiently bind with repressor without being metabolized by *E. coli*, therefore its quantity remains constant during the expression interval (figure: 1.9.5.B).

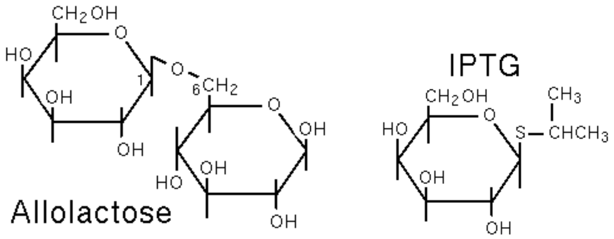


Figure 1.9.5.B: Lac Operon inducer Allolactose and homolog IPTG.

The greater transcription rate is only observed when glucose is depleted in the media. Depletion of glucose means, increased level of cAMP, that forms a complex with CAP (catabolic activator protein) and binds with the CAP binding region situated above the promoter region and facilitate the RNA polymerase binding to the promoter region.

Based on the impact of glucose on effective translation, an **auto induction system** has been developed based on the balance between the glucose and lactose in media. Studier et al. has done intensive experimental studies with the components of defined media (ZYM-505) and conditions (Temperature, timing, pH and aeration). Their studies have concluded in media composition and culture conditions best suited for automatic conversion of cell higher density toward cloned gene expression conditions.

This expression is dependent on glucose depletion from the media and resulted cAMP increase in media to allow high level expression (Studier 2005). This is very efficient and intelligent technique for automated induction of cells without IPTG induction, but do not seem well studied for toxic protein expression as there is always leaky expression present before the complete depletion of glucose from the media (figure: 1.9.5.C).

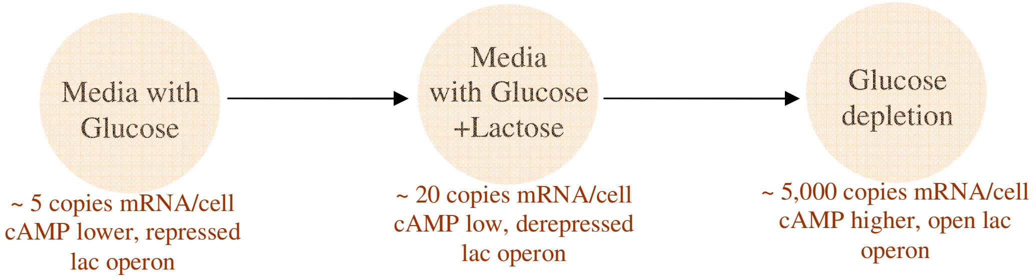

Media with Glucose
~ 5 copies mRNA/cell cAMP lower, repressed lac operon
Media with Glucose +Lactose
~ 20 copies mRNA/cell cAMP low, derepressed lac operon
Glucose depletion
~ 5,000 copies mRNA/cell cAMP higher, open lac operon

Figure 1.9.5.C: Mechanism of expression in auto induction media used with *lac* operon systems.

II. *pBAD* or *AraC* Operons:

This promoter is frequently use in cloning systems for tight regulation of basal expression such as in pBAD/gIII vector (invitrogen®) for gene expression, in BL21-AI cells for T7 RNA polymerase synthesis etc. The *ara*BAD promoter transcriptional regulation is both positively and negatively regulated by *ara*C gene product. In the absence of L-arabinose the AraC dimer contacts the O2 and I1 of the *ara*BAD operon, as a result the intervening DNA form a 210bp loop of araC coding site. This confirmation leaves no access for RNA polymerase, to start transcription from $\mathbf{P_{BAD}}$ promoter, hence there is a negative regulation by araC in the absence of L-arabinose (Hirsh and Schleif 1977).

When arabinose is present, it binds to araC and allosterically stimulates the release of O2 site. So, that DNA loop formation resolves, leaving the pBAD promoter open for the access of RNA polymerase for active transcription. This positive regulation needs the assistance of CAP-cAMP complex to bind with CAP binding site and brings the right conformation for the I1 and I2 bridging by allosterically modified araC diamer (Lee, Francklyn et al. 1987). In the absence of glucose the cAMP level is higher in cell and CAP-cAMP complexation occurs, but when glucose provided in media, cAMP involves in the glucose catabolism pathway and no CAP-cAMP complex forms. This is called glucose repression of expression (Lee, Wilcox et al. 1974).

AraC also play a role for autoregulator of its own expression in the DNA loop confirmation. When araC scares no loop can be formed and transcription starts form the Pc promoter for araC expression. During the scars of araC, pBAD promoter can also be transcribed resulting in a leaky expression (Irr and Englesberg 1971).

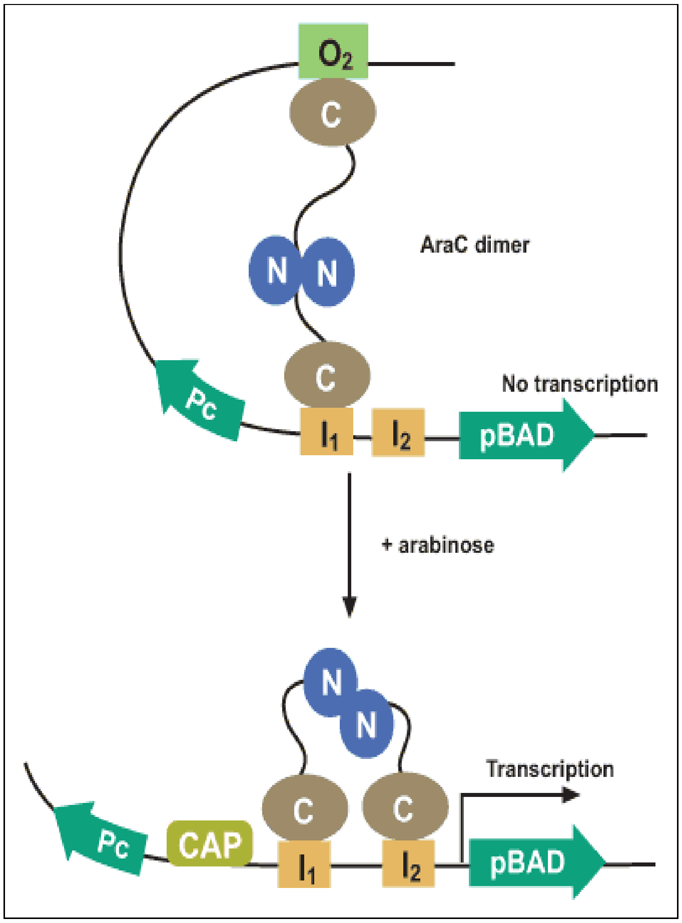


Figure1.9.5.D: explanatory diagram for the araC Operon regulation with and without arabinose. Where Pc is promoter regions of araC and pBAD is promoter for arabinose catabolic structural genes. Source; commerce invitrogen.com

1.9.6. Plasmid copy number/ origin of replication:

Plasmid copy number depends on the type of origin of replication present in the vector. Most of the today's vectors origins are mutated from few known types to convert them in high copy numbers (Schmidt, Friehs et al. 1996). Examples of some of the high and low copy number plasmids are given in following table.

Table 1.9.7: plasmid origin and plasmid copy numbers.

Prototype plasmid	Derived /originate from	Size (bp)	Selectivity Marker	Gene dosage	Category (copy number)	Reference
pTZ	(pMB1)	2,862	Amp	>1000	Very high	-
pUC 18/19	(pMB1)	2,686	Cam	500-700	High	(Lin-Chao, Chen et al. 1992)
pBluescript	(ColE1)	2,961	Amp	300-500	High	(Mayer 1995)
pGEM	(pMB1)	3,000	Amp	300-400	High	(Young, Matsubara et al. 1996)
pBR345	(pMB1)	1,100	Cam	100-300	High	(Bolivar, Betlach et al. 1977; Bolivar, Rodriguez et al. 1977)
pMK16	ColE1	4,500	Kan, Tet	>15	high	(Kahn, Kolter et al. 1979)
pBR322	(pMB1)	4,362	Amp, Tet	100-200	high	(Bolivar, Rodriguez et al. 1977; Bolivar, Rodriguez et al. 1977)
pACYC 184	(p15A)	4,000	Eml, Tet	~15	Low	(Chang and Cohen 1978)
pLG338	(pSC101)	7,300	Kan, Tet	~5	Very low	(Stoker, Fairweather et al. 1982)
pDF41	(F)	12,800	TrpE	1-2	Very low	(Kahn, Kolter et al. 1979)
pBAC/oriV	(*oriV*-TrfA)	8,000	Amp, Cam	~1 (un induce)/ 100x (induced)	Single (SC)/ High (HC)	(Wild and Szybalski 2004)
pBEU50	R1 (R1*drd*-17)	10,000	Amp, Tet	-	Low-30ºC/ high 35ºC	(Uhlin, Schweickart et al. 1983)

In recent advances, replication of plasmid has also been tried to restrict when cell is not induced so that leaky expression in uninduce cell can be tightly regulated. This novel replication origin is under its own control, and remains as single copy when cells are uninduced. As soon as inducer added to the media, active replication of plasmid starts and reaches 100 folds (Wild and Szybalski 2004). Similar approaches have also been in application with rhamnose-inducible *Prha* promoter (*rhaS-Prha*). In expression systems such *PLtetO-1* as promoter and *pT7lacO* promoter in pETcoco vector series, an independent regulation carried out for both plasmid amplification and cloned gene regulation (Wild and Szybalski 2004).

1.9.7. Applications of fusion tags:

On of the main factors in the choice of expression vector is the need for fusion tag/ tags, for **facilitation of purification**. The most common is the His tag which is relatively small tag, with often minor effect on the overall structure and do not interfere much in protein structure determination using X-ray crystallography or

NMR, in most of the cases. Other purification assisting fusion tags includes, FLAG, IMPACT, G protein, GST, *c-myc,* MBP, T7, S, strep, CBD, DHFR tags. Among them FLAG (SIGMA-aldrich) which is very small in size, is a multipurpose tag because it can be use for detection through ELISA or western blot, can use in affinity purification, and it is target for Enterokinase cleavage site. Therefore, it can be removed after purification with affinity chromatography. IMPACT (Intein-Mediated Purification with an Affinity Chitin-binding Tag) is also another advantageous system, which can be purified using affinity to chitin column. On the other hand this accessory protein can be cleaved off by self splicing through intine (protein splicing element) in reducing environment. Other tag systems like GST and His are now been facilitate with rapid **purification spine columns** tubes and **pull-down assays.**

Protein tags can also provide assistance in **solubility** of proteins. Like thioredoxin, Z tag , NusA, MBP, Gb1, G-protein are well known for increasing the solubility of insolubility tending proteins (Hammarstrom, Hellgren et al. 2002). **Detection** is another purpose of expressing protein with Tag. Some proteins express in small quantity and can not be seen on the SDS-PAGE. Then it needs to be detected in initial stage of the expression. EGFP, HSV, *c-myc,* V5 Epitope, FLAG, His, T7, CAT, CBD, DHFR tags are those that can be detected for expression. EGFP is exceptional for **real time detection** of signal protein travel, inside the cell. HSV tag can also be use for immunohistological assays.

Some tags are also used for the **quantitative estimation** of expressed proteins like the GST tag, wich detected on the base of GST enzymatic activity. Some tags associate with the **small proteins/ peptide expression** like Gb1, Ubiquitin, DHFR, and KSI tags. There are some accessory proteins that use to **transport the expressed protein** in the E. coli periplasm or for secretion in other cell systems. These accessory tags example include GIII, PelB/ompT, DsbA, DsbC, $\mathrm{CBD}_{\mathrm{cenA,}}$ $\mathrm{CBD}_{\mathrm{cex.,}}$ among them GIII, PelB/ompT act as transporter proteins only. They are cleaved off from target protein after the transport to the destination. While DsbA, DsbC, $\mathrm{CBD}_{\mathrm{cenA,}}$ $\mathrm{CBD}_{\mathrm{cex}}$ remain attached after the transport of target protein. The previous two (DsbA, DsbC) are used to perform the disulphide isomerization function, the further two ($\mathrm{CBD}_{\mathrm{cenA,}}$ $\mathrm{CBD}_{\mathrm{cex}}$) are used to perform the assistance in purification and detection.

Table 1.9.7: Fusion tags provided with vectors for purification, detection, quantitative assays and solubility

Fusion Tags/ insertions	N/C - terminal/ I	Size (a.a)/ Weight (KD)	source	Basis of detection/ traits	Application	Reference
GIII (pIII)	N	18/-		Periplasmic localization	Export of protein to periplasm	invitrogen
pelB/ompT	N	20, 22/-		Periplasmic localization	Protein export, folding, partial purification	Novagen
DsbA	N	208/-		Periplasmic localization	Solubility, Periplasmic disulphide formation, isomerization	Novagen
DsbC	N	236/-		Periplasmic localization	Solubility, Periplasmic disulphide formation, isomerization	Novagen
c-myc	C	10/ 1.2		*Myc* Epitop	Anti-Myc /Myc-HRP antibodies, affinity purification	invitrogen
His	N, C, I	6, 8, 10/-		Antibodies/ affinity chromatography	Western blot, IMAC purification	Novagen
Z-tag	N	58/ 7		Z-domain (derived from protein A)	Enhanced Solubility, Anionic chromatography	SPINE (Stockholm)
Gb1	N	18.6/ 56		Good stability	Small peptide expression, solubility	SPINE (Stockholm)
GST	N	56/ 26.2		Monoclonal antibodies, Glutathione affinity, Enzymatic activity	Western blot, purification, quantitative assay	Novagen
G protein	N	68/ 8.2		IgG affinity	Affinity Purification	Not commercial
MBP	N, C	400/ 44.2		Amylose affinity	Solubility enhancement, affinity purification	NEB
NusA	N	495/ 56.4		Monoclonal antibody	Solubility enhancement	Novagen
Trx	N	109/ 11.8		Monoclonal antibody, ROS-scavenging function	Soluble protein, disulphide reduction.	Hamburg
Ubiquitin	N	76/ 8		Good stability	Peptide expression, solubility	-
T7	N, I	11/ -		Monoclonal antibody	Western blot, immunopreci-pitation, affinity purification	Novagen
S	N, I	15/ -		S-protein affinity, enzymatic activity	Western blot, quantitative assay, affinity purification	Novagen®
DHFR	N, C	190/ 22		DHFR is non immunogenic, methotrexate affinity	Antibody preparation, Expression of peptides and small proteins, purification	(Vermersch, Klass et al. 1986)
EGFP	N, C	238/ 26.9		Green florescent protein	Cell signaling studies, protein localization studies	BioVision
CAT	N	-/ 24		Chloramphenicol-O-Acetyltransferase	Blotting, fluorometric assay, transcription assays	-
Strep	N	8		Monoclonal Antibody	Strep.Tactin affinity purification	Novagen
HSV	C	11		Monoclonal antibody	Western blot, immunoflorescence	Novagen
KSI	N	125/ 13.5		Highly expressed hydrophobic domain	Small proteins/ peptide expression / purification	Novagen
CBD$_{clos}$	N	156/ 17		Polyclonal antibody/ cellulose binding domain	Western blot, purification, non covalent immobilization	Novagen
CBD$_{cenA}$	N	114		Polyclonal antibody/ periplasm / media, cellulose binding domain	Western blot, purification, non covalent immobilization, protein export	Novagen
CBD$_{cex}$	C	107/ 10.8		Polyclonal antibody/ periplasm / media, cellulose binding domain	Western blot, purification, non covalent immobilization, protein export	Novagen
IMPACT	N, C	500/ 60		Chitin affinity of protein, splicing element (*Intein*)	Chitin affinity purification, *Intein* mediate, tag self splicing	NE BioLabs

1.9.8. Special function sites in vectors:

Attachment of very large solubility tags usually interfere with the structural, activity and in the antigen production and immunogenic reactions. Therefore they need to be cleaved off from the target protein afterwards. Special cleavage sites are now been provided in between the target protein and the tags for removal of tag protein. These sites are listed in the table 1.9.8. In addition to cleavage sites special phosphorylation sites are in some instance of special interest for isotopic labeling, quantitative estimation of target protein etc.

Table 1.9.8: Common sites provided with in vectors for special functions.

Special sites	Positions	a.a.	sequence	Function
Enterokinase cleavage site (FLAG-Tag)	Between tag-target	8	DYK ↓DDDDK	Tag removal, detection ELISA, western blotting, affinity purification
Factor Xa cleavage site	Between tag-target	4	IQ/EGR↓	Tag removal
Thrombin cleavage site	Between tag-target	6	LVPR↓GS, [BamHI]	Tag removal
PreScission cleavage site	Between tag-target	8	LEVLFQ↓GP, [ApaI]	Tag removal
TEV cleavage site	Between tag-target	7	ENLYFQ↓G/S, [StyI], [BamHI]	Tag removal
HRV 3C cleavage site	Between tag-target	8	LQVLFQ↓GP	Tag removal
Genenase 1 cleavage site	Between tag-target	6	PGAAH↓Y↓	Tag removal
Caspase-3	Between tag-target		DXXD↓	Tag removal
pKA site	N- terminal	5	RRASV	*In vitro* Phosphorylation, Isotopic labeling

1.9.9. Choice of expression host for quality and quantity increase in protein expression

Expression cells are selected on the bases of their suitability for target protein and special advantages integrated with the cell type. Commonly there are five types of expression systems known, that are E. coli, Yeast, insects cells, mammalian cells and cell lysate (from E. coli, wheat germ cells, etc.). Except them plants, plant cells, fungus (*Aspergillus niger, A. Oryza, Fusarium venenatum*) and gram positive bacteria (*Bacillus subtilis SK-52*) are reported in expression with some advantages over commonly known systems. In industrial enzyme production, *Bacillus spp*. has taken little more attention over traditional gram negative prokaryotic host *E. coli*, due to its efficient secretion system. Since this organism is less studied in comparision to E. coli, therefore some compensation need to be still made to make this organism ideal for industrial enzyme production. These compensations are essential for protease deficient strain development and for improving the secretion machinery for over expressed protein transport (Simonen and Palva 1993).

Table 1.9.9: Characteristics of different protein expression systems.

Characteristics	*E. coli*	Yeast	Insect cells	Mammalian cells	cell free system (Roch RTS)
Example	BL21, lysogenic strains (DE3)	*Pichia pastoris, Saccharomyces cerevisiae*	SF-9, SF-21, ovarian cells of Fall Armyworm	GS-CHO, Hybridomas	*E. coli lysate*, Wheat germ lysate
Proteolytic cleavage	yes	yes	good	good	No
O-linked Glycosylation	No	yes	yes	good	No
N-linked Glycosylation	No	Mannose required	yes	good	??
Phosphorylation	No	yes	yes	good	No
Acetylation	No	good	yes	good	No
Acylation	No	good	good	good	No
Amidation	No	good	good	good	No
γ-Carboxylation	No	No	No	yes	No
Secretion	Periplasm yes, media not good	Secretion to media good	Secretion to media good	Secretion to media good	N/A
folding	lacking	likely	proper	proper	proper
Media (cost)	Rich (low)	Rich (low)	Complex (medium)	Complex (high)	Complex (higher)
Cell growth (hrs)	Rapid (3-10)	Rapid (5-16)	Slow (18-24)	Slow (18-24)	Very fast
Yields mg/L	50-500	10-200	10-300	0.1-100	100-500/hrs
Advantages	Simple, robust, plenty of options	Simple, better PTM & secretion	Good folding & PTM, Higher yield	Native environment for HG in folding and PTM	Real time studies possible
Pitfalls	No PTM, inefficient in secretion	Lesser options, longer time	Longer time, higher cost	Higher cost, lowest yield	No PTM, Highest cost
recommended	Industrial expression, Structural & functional studies, antigen production	Industrial expression, Structural & functional studies, vaccine production	Structural & functional studies, PTMs, standard assays	Eukaryotic proteins' Cell signaling studies in native environment, bioassays.	Functional studies, real time experimentation, labeling for NMR structures
Project cost	lowest	lower	Middle	higher	higher

While; PTM stands for post translational modification (++) stands for very good, (+) stands for not so good and (-) stands for not good. (*source:(Labaere, Gruenwald-Janho et al. 2004)*

Expressions in mammalian cells are although expensive and labor intensive, but some time it can not be avoided. Since, it presents a significantly important source for studying cell signaling pathways, in drug target discovery and validation, physiological and pathological studies and in gene therapeutical studies under *ex vivo,* native conditions (Barnes and Dickson 2006; Chartrain and Chu 2008).

The most recent among all is the cell free system that is still in developing stage. In advancement of this system many type of cell lysate have been introduced and are under investigation. Until now *E. coli cells'*, *wheat germ cells'*, *mammalian cells'* (rabbit reticulocyte), *incest cells'* lysates have been reported for *in-vitro* protein synthesis application. The advantages of this laborious and expensive technique are higher in some areas like membrane protein expression in hydrophobic environment and efficient multi radiolabelling for NMR studies (Staunton, Schlinkert et al. 2006; Schwarz, Dotsch et al. 2008).

1.9.10. Expression in *E. coli*:

E. coli is the most popular cell on laboratory scale and to some extent in industrial preparation of enzymes. But it is not much in use for the commercial scale preparation because it is poor in secretion system and it lacks most of the posttranslational modification apparatus resulting in poor folding and non functional proteins. But still it is popular because it is the well studied organism, easy to grow and manage to the high cell densities within hours. Since it is most studied organism transformation and cloning protocols are well developed. Until present, most of the plasmid vectors have been developed for *E. coli* systems with advancement toward tight regulation of gene expression and enhanced productivity (Greene 2004).

In recent progression there are several modified cell lines available that contain additional chaperones to assist the proper folding (e.g. **BL21(DE3) GroES/L**) (Caspers, Stieger et al. 1994; Luo and Hua 1998; Endo, Tomimoto et al. 2006). Some cells contain rear codons to help external eukaryotic proteins to express better (e.g. **B834 (DE3) pRARE, BL21 (DE3) CodonPlus**, **Rosetta** series). Some of the cells modified with the mutation in the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes that assist the disulphide bonding in cytoplasm for proteins that require cystine briges in their structure for proper folding (e.g. **AD494 (DE3)**,

Origami series). It is reported that use of these kind of mutated cells can enhance 10 folds increase in the activity of the soluble proteins (Levy, Weiss et al. 2001; Cassland, Larsson et al. 2004; Saejung, Puttikhunt et al. 2006). Some bacterial strains lack proteases to control over protein degradation in host like, **BB7333** lacks *clpX, clpP* and *lon* proteases, **BL21** series lacks *lon* and *ompT.* For some strains increased stability of the protein has been tried to be achieve on mRNA level, such as in **BL21 Star** strains. **B834 (DE3)**, **DL41 (DE3)** are methionine auxotroph cells that permits the labeling of proteins with ^{35}S-methionine and selenomethionine, for autoradiography and X-ray crystallography (Fuss and Godwin 1975).

pLysS and **pLysE** are modified for toxic protein expression, as they carries a T7 lysozyme encoding gene (a natural inhibitor of T7 RNA polymerase), that can efficiently inactivate the leaky expression of RNA polymerase. pLysS produce lesser amount of T7 lysozyme, while pLysE produce much more enzyme and is therefore more stringent (Dubendorff and Studier 1991; Studier 1991). **BL21 AI**, **LMG194**, has tight regulation with araBAD promoter system hence efficiently control over leaky expression. High Stringency **T7lac Promoter** in Tuner (DE3) **pLacI** is also an option to control the basal expression for toxic protein expression. **Tuners** are lactose permease (*lacY*) mutant that enables uniform uptake of IPTG hence enable full force utilization of inducer (IPTG), in induction of operator.

1.10.0: problems in protein expressions:

In following sections, problems related with protein expression will be discussed in detail, among all problems insolubility is the most invasive phenomenon that results in inactive protein and difficulties in purification.

1.10.1. AT rich genome:

AT rich genome from *Plasmodium falciparum* had found extremely difficult to express in prokaryotic and eukaryotic expression system, due to more abundant rear codons, and in *Pichia pastoris* due to the rich AT based polyadenylation or transcription termination signals. The genetic mutation that can lead to codon abundance was found as a solution to the problem (Withers-Martinez, Carpenter et al.

1999). In case of *E. coli* overlapping leader open reading frame was found the solution to expression of AT rich genes (Ishida, Oshima et al. 2002).

1.10.2. Initiation codons and sequences before it:

Poly purine domain UAAGGAGGU known as the Shine-Dalgarno (SD) is essentially important for the recognition of translation initiation site on mRNA. Spacing between the SD and the initiation codon strongly affects translational efficiencies (Chen, Bjerknes et al. 1994; Makrides 1996). At least 8 base pair A+T rich translational spacer should be present between the initiation codons AUG (M), GUG (V), UUG (L), that have respectively 91%, 8% and 1% translational efficiency in E. coli (Makrides 1996). A translational enhancer sequence from T7 phage called g10-L has also been discovered to enhance the translational efficiency 40-300 fold (Olins and Rangwala 1989).

1.10.3. Presence of rear codon on gene:

The frequency of codon usage varies from organism to organism. Even for an organism, nature has controlled the specialized genes expression sometime by inserting rear codon containing genes in the sequence (Chen, Bjerknes et al. 1994; Saier 1995). If a gene from an organism will try to express into another organism, then the difference in codon usage frequency from host organism to the cloned gene's species can cause problems during over expression of cloned gene. These problems include, decrease in expression efficiency (if rear codon/codons are in the starting of sequence), frameshift mutation +1/-1, mismatch pairing of mRNA codon and antisense tRNA, (Sharp, Stenico et al. 1993). Clustering of rear codons in genes leads to very poor expression or frame shift mutation (Kane 1995). Table 1.10.3; describe frequencies of reare codons in *E. coli*

There could be two possible solutions to rear codon occurrence in a desirable gene expression. An intelligent solution was presented by Anson and dunning have mutationally converted the rare codons from HIV-1 minor proteins' into abundant codon of host organism, with obvious increment in yield (Anson and Dunning 2005). The other solution to rare codon is the co- expression of rear codon in host species, so that it would not become scares during expression (Fu, Lin et al. 2008). Several rear codon expressing strains are available for this purpose like Rosetta (Novagen), that harbor additional plasmid to encodes AGA (Arg), AGG (Arg), CGA (Arg), GGA (Gly), AUA (Ile), CUA (Leu), CCC (Pro); BL-21 codon plus-RP (stratagene), that harbor plasmid encodes only CCC (Pro), AGA (Arg), AGG (Arg) and BL-21 codon plus-RIL (stratagene), that express AGA (Arg), AGG (Arg), AUA (Ile), CUA (Leu).

Table 1.10.3: E. coli's rare codons [source: (Saier 1995)]

Rare codons	Encoding a.a	Frequency/1000 codon
UAG	**Non sense**	0.3
UGA	Non sense	1.0
AGG	**Arg**	1.4
UAA	Non sense	2.0
AGA	**Arg**	2.1
CGA	**Arg**	3.1
CUA	**Leu**	3.2
AUA	**Ile**	4.1
CCC	**Pro**	4.3
CGG	**Arg**	4.6
UGU	**Cys**	4.7
UGC	**Cys**	6.1
ACA	**Thr**	6.5
CCU	**Pro**	6.6
UCA	**Ser**	6.8
GGA	Gly	7.0
AGU	Ser	7.2
UCG	Ser	8.0
CCA	Pro	8.2
UCC	Ser	9.4
GGG	Gly	9.7
CUC	Leu	9.9

Irrespective of all the solutions of rare codon scarcity, codons universality for amino acid coding is not universal in special organs of metazoan phylogenetic tree. It should be noted that there have been modifications in five of the sixty-four codons i.e. TGA (stop), ATA (Ile), AGA (Arg), AGG (Arg), AAA (Lys) (Wolstenholme 1992). In mammalian mitochondria TGA codes for Trp instead of stop codon; ATA codes for Met instead of Ile; AGA and AGG codes for Stop codon instead of Arg. These exceptions can be seen in all type of mitochondria from metazoan (Mammalians, Nematoda, Echinodermata, Cnidaria, Platyhelmenthese etc.) (Osawa, Ohama et al. 1989; Osawa, Ohama et al. 1989)

1.10.4. Size of Gene:

As the gene size increases, tendency of common expression problem tends to rise. These problem include no expression, insolubility, truncation etc (Dyson, Shadbolt et al. 2004; Luan, Qiu et al. 2004). If a plasmid becomes extremely heavier due to the large gene insert, then high copy number tends to shifts toward low copy number (Novagen 2003).

1.10.5. N-end rule:

The half life of the proteins in mammalian, yeast and bacterial system, determined on the basis of their *N*-terminal residue (Bartel, Wunning et al. 1990; Tobias, Shrader et al. 1991). It has been experimentally explored with the different *N*-ends, engineered and examined with Beta-galactosidase in the models of yeast in vivo; mammalian reticulocytes in vitro, Escherichia coli in vivo systems (Gonda, Bachmair et al. 1989; Bartel, Wunning et al. 1990; Tobias, Shrader et al. 1991). Figure 1.10.5 summrizes the role of different amino acids in the *N*-terminal of expressed proteins that have major or minor attraction to protease degradablity.

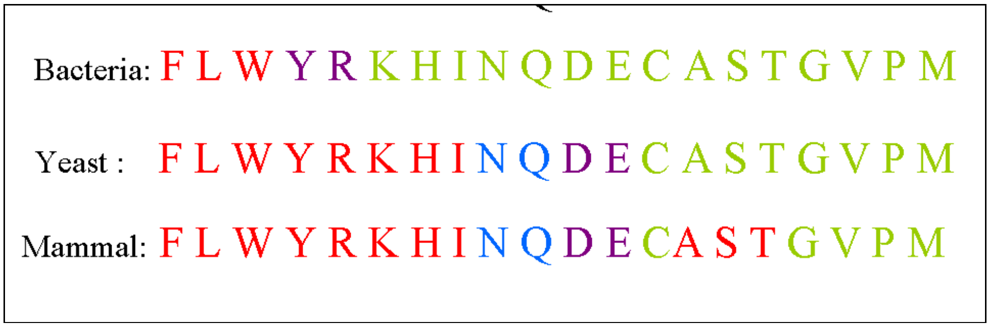

Bacteria: F L W Y R K H I N Q D E C A S T G V P M
Yeast : F L W Y R K H I N Q D E C A S T G V P M
Mammal: F L W Y R K H I N Q D E C A S T G V P M

Figure 1.10.5: Amino acid code shown with Primary (red), secondary (purple), tertiary (blue), de-stability residues in *N*-end rule. Source: (Tobias, Shrader et al. 1991)

1.10.6. Instability of mRNA:

Several stabilizing and degrading factors are now known for the control on the fate of mRNA in E.coli. In degrading factors endonucleases (RNase E, RNase K, and

RNase III) and 3'-exonucleases are known to play role. In stabilizing factors UTRs of mRNAs and stem-loop structures are known to stabilize mRNA (Makrides 1996). Except that some efforts have made toward development of strains that lacks RNAase E like in BL21 Star (DE3) cells.

1.10.7. Plasmid loss:

When there is insufficient amount of antibiotics or plasmid is of 'low copy number' then usually, 'plasmid less culture' develops that gives the false negative impression of little or no expression. Ampicillin resistant culture grown for longer period of time also exhibit this kind of phenomenon, since with increasing time, beta lactamase secretion increases in media and destroy beta lactum of ampicillin.

Culture grown at lower temperature and addition of glucose usually enhance the stability of plasmid (Zhang, Taiming et al. 2003).

1.10.8. Plasmid copy number:

The yield of cloned target also vary with the plasmid copy number, generally it is thought that high copy number results in overexpressed target protein (Friehs 2004). But in some applications like solubility enhancement and metabolic engeneering, low plasmid copy number (LC) are preferred over high copy number plasmids (HC) (Kim and Keasling 2001; Mergulhao, Monteiro et al. 2003). Exceptional to their general copy number information, some other factors can influence the copy numbers like growth conditions and size of cloned genes (Feinbaum 2001).

In recent advances, replication of plasmid has also been tried to restrict when cell is under non expressional condition, so that leaky expression can be controled tightly. This kind of novel replication origin is under its own control, and remains as single copy when cells are uninduced. As soon as inducer added into the media active replication of plasmid starts and reaches 100 folds (Wild and Szybalski 2004). Similar approaches have also been in application with rhamnose-inducible *Prha* promoter (*rhaS-Prha*). In expression systems such as *PLtetO-1* promoter and *pT7lacO* promoter in pETcoco vector series, an independent regulation carried out for both plasmid amplification and cloned gene regulation (Wild and Szybalski 2004).

1.10.9. Disulphide bond formation:

Proteins that need disulphide bonds for their proper folding or oligomerization can not be expressed inside the reducing environment of *E. coli* cytoplasm. On expression inside the cytoplasm they either degrade due to the improper folding or go into the inclusion bodies in inactive form. There are three alternative to this problem; (i) Expression of target proteins in the hosts deficient in *thioredoxin reductase* & *glutathione reductase* (enzymes responsible for reduction of disulphide bonds). One of these type is AD494(DE3) strain (Novagen), which has single mutation for *thioredoxin reductase gene,* and enables the expression of C1 inhibitor, containing two disulphide bridge (Simonovic and Patston 2000).

Another more strongly acting modified strain is Origami (Novagen), which has double mutation for *thioredoxin reductase* & *glutathione reductase genes*. (ii) The second option is to add a functional signal sequence adapted to the host system, in order to translocate the target protein toward the periplasm. The periplasm has reducing envoirnment and enzyme machinery for proper disulphide bridging. (iii) The third option is to express the protein with fusion tags DsbG, DsbC that are well known for their chaperone activity and protein disulphide bond isomerization (PDI) (Andersen, Matthey-Dupraz et al. 1997; Chen, Song et al. 1999; Zhang, Li et al. 2002). One example is with DREBIII-1, a plant specific transcriptional factor, that has markedly increased the activity as well as solubility when expressed with PDIs (Protein Disulphide Isomerase) (Liu, Zhao et al. 2005).

1.10.10. Choice of *N/C*-terminal tags:

Decision to select the site of tag attachment, dependent on the exposure of the terminal on the surface of protein. Therefore, it is important to reveal the exposure of terminals onto the surface of protein, before making the decision of position of attachment. Exposure of terminals can be revel with the knowledge of known structure, or with the model building for unknown structures.

If both of the terminals are exposed, tags can be attached on both terminals (e.g. pBAD-DEST49, *N*-terminal HP-thioredoxin, C-terminal V5-6xHis), to obtain the diversity in detection and purification procedure of proteins. If one of the terminals or both of them are unexposed to protein surface, then attachment of tag in buried surface can be resulted in expression leaded to the inability of native folding or degradation (Klose, Wendt et al. 2004).

Attachment of accessory protein in buried surface can also resulted in inactive protein due to the misfolded conformation in protein active sites (Goel, Colcher et al. 2000). When *N*-terminal is exposing, poorly express protein can be up regulated if accessory protein has solubility enhancing capabilities. Position and size of tag both effect differently from protein to protein (Mohanty and Wiener 2004).

Mostly, *N*-terminal tag is preferred over *C*-terminal tag, because in some cases *C*-terminal tags are reported to interfere in the folding of protein and consequently to poor or undetectable expression (Dyson, Shadbolt et al. 2004). In case of exposed *C*-terminal, detection tag can ensure the complete translation of the polypeptide. When the structure of a cloned protein is not known, attempts can be made for both terminal and construct having a good solubility and maximal activity can give the evidence of proper native folding (Sachdev and Chirgwin 1998).

1.10.11. Selectivity marker:

Enhanced accumulations of drug resistance RNA are reported to deposite upon induction, when selectivity marker (antibiotic resistance gene) orient, in same diraction as target gene. The phenomena is said to happen due to the readthrough over the T7 transcription terminator that is only 70% efficient upon induction. One of the solution of this problem has been introduced by Novagen in kan^R pET vectors, where they have kept ‘selectivity marker gene’ in the opposite direction to target gene (Novagen 2003). This option has not only resolve the over production of *β–Lactamase* in induce culture but also solve the problem of selectivity loss at the end of the expression stage due to the increased degradation of β–Lactam (ampicillin).

1.10.12. Truncated proteins: (Mutation/ secondary initiation site)

Sometimes truncated protein can be seen in the expression. The reason could be the possible mutation during the cloning procedure by over exposure of DNA with UV or by growing the cells at higher temperature. If the final clones have been sequenced and there will be no mutation in the form of stop codon, then another possibility could be the presence of a secondary translation site within the cloned sequence. Secondary translation site is a similar sequence to ribosome binding site, that can be 5-13 nucleotide upstream with ATG (initiation Methyl codon) (Novagen 2003).

1.10.13. Physiological conditions:

For a decided expression system (combination of expression vector and expression cells) alteration of physiological conditions and media composition can bring drastic changes into the protein expression, quantitatively and qualitatively. Such factors include temperature, salt concentration, pH of media, inducer concentration, media constituents etc.

1.10.14. Protease degradation susceptibility:

Several protease recognition motives are studied that are believed to act universally. Except the *N*-end rule, proteins that have nonpolar amino acid in their *C*-terminal is believe to undergoes rapid degradation, while proteins that have charged or polar amino acid fails to degrade (Makrides 1996). Care should be taken according to these established rules but exposure of *N*-or *C*- terminal is another important factor that governs the ability of these rules. Among other susceptibility factors, cultures conditions like optimal temperature and pH for proteolytic activities are also important to consider. Incomplete, damaged or instable protein can also rapidly undergoes proteolysis.

1.10.15. Instability of protein:

Purified and concentrated solutions usually show the instability. They suffer from aggregation and precipitation problems. Screening for proper buffer condition is one of the solutions that is a tedious job. Additions of stabilizers are also known to work in this condition like L-Arg and L-Glu in 50mM concentration dramatically stabilize the concentrated protein solutions for longer term. Moreover protections of protein degradation by proteolytic enzymes are also reported using this method.

(Golovanov, Hautbergue et al. 2004). Other additives include osmolytes, glycine and detergents.

1.10.16. Problems with cold adapted protein Expression*:*

Cold adapted proteins are usually temperature sensitive and exhibit lower stability at higher temperature. Except their vulnerability with higher temperature they are also unstable at lower temperature, particularly in the active site, when compare to their thermostable homologs. This is consistent with their less packed hydrophobic core and fewer ionic interactions (Siddiqui and Cavicchioli 2006). Therefore enzymes from cold adapted sources are much more vulnerable to destabilization related precipitation and degradation.

Cold adapted proteins are predicted to contains more hydrophobic surface residues in comparison to their mesophilic and thermophilic counter parts (Thorvaldsen, Hjerde et al. 2007). Therefore they often face problems in purification due to the soluble aggregate formation.

1.10.17. Insolubility (Aggregation / Inclusion bodies/ Misfolding):

Expression in *E. coli*, some time leads to aggregates or inclusion bodies formation. These aggregates some time contain denatured, misfolded and non functional target protein and some times contain the material other than the target proteins like bacterial carbohydrates etc. Protein folding is a complex phenomenon where several factors affect the phenomenon and leads to either proper folded, misfolded (non-functional) or jumbled (aggregated) proteins. These misfolded and jumbled proteins can some time remains soluble and appear as non-functional and soluble aggregates. Properly folded proteins can also some time remains insoluble by forming aggregates due to their surface properties and mainly due to the improper buffer conditions.

Factors that effect the proper folding of proteins irrespective of their compositional, and structural characteristics are temperature, pH and salt (Cabrita and Bottomley 2004). These factors can also affect the solubility of properly folded proteins. Incompatible variation in these factors can lead to the precipitation, soluble aggregate formation and some time unfolding of the properly folded proteins (figure: 1.10.17 A, B, C).

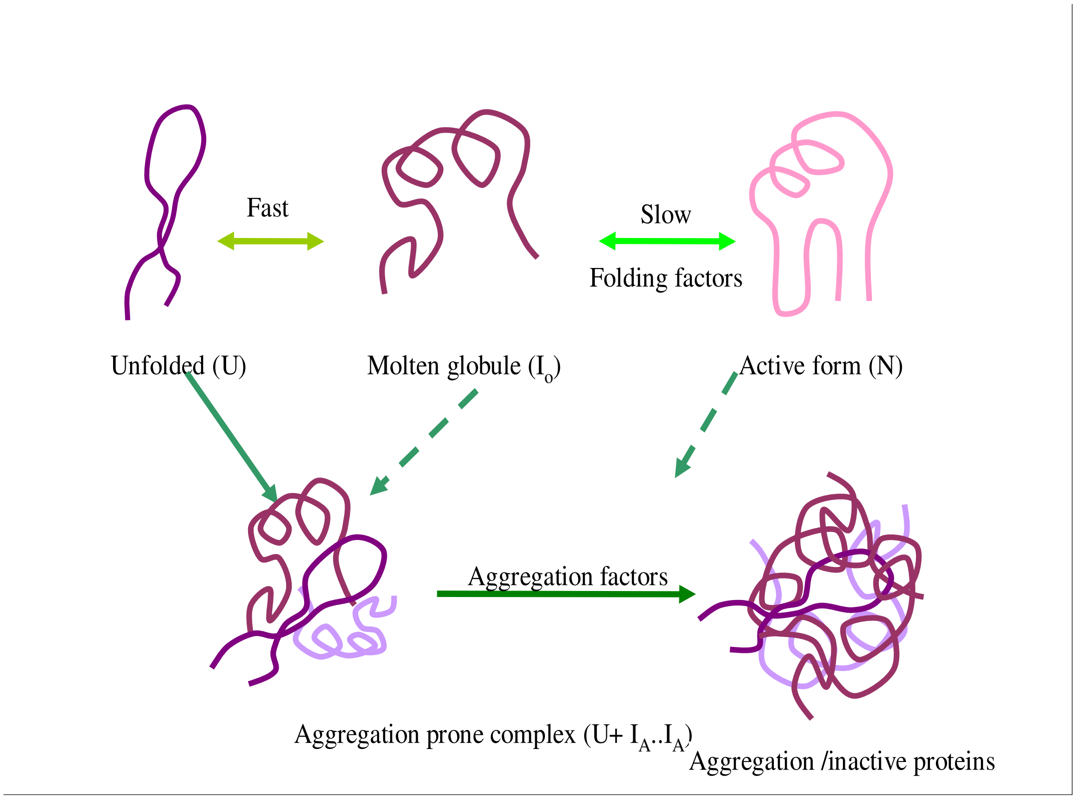


Figure 1.10.17.A: mechanism of folding and aggregation. While, I_0 is intermediate stage in folding, N is native folded state, I_A is aggregation prone intermediate. Temperature, pH and salt are folding factors. Aggregation factors depends on amino acid composition of each individual protein [derived from (Cabrita and Bottomley 2004)].

The Harrison group have compiled the data from the 81 different soluble and insoluble proteins in Oklahoma University, and tried to correlate them with their amino acid sequences. They have found out six exclusive parameters that influence the inclusion body formation. They ordered with severity from average charge, fraction of turn forming residue, fraction of cysteine, fraction of proline, hydrophilicity and the total number of residues (Wilkinson and Harrison 1991).

Folding is a process where two main factors influence on the correctness and rate, these are correct cysteine bond formation, and turn forming rate. Proline, asparagine, glycine, serine and aspartic acid are amino acids that often occur in turns. Among them, folding of proline has been found most crucial in the rate limiting for turn formation in several proteins folding (Evans, Dobson et al. 1987).

Protein solubility is directly proportional to net surface charge of proteins (negative or positive). With in low ionic content solutions, log of solubility is directly proportional to the square of net charge of proteins (Tanford 1958).

Log [protein Solubility] = (protein net charge) ^{2}…….*Debye Hückel equation*

While at neutral pH

$$\text{Net charge} = \frac{\text{Average charge}}{\text{Residue}} = \frac{(\text{Basic residues} - \text{Acidic residues})}{\text{Total number of residues}}$$

Debye Hückel equation also describes the dependency of protein size on the solubility. Size of protein can be roughly estimated with the number of residues in proteins.

Hydrophobicity on protein surface is the important factor that initiates the aggregation or inclusion body formation. These aggregates or inclusion bodies remain insoluble in the absence of inappropriate buffer conditions and sediment on the process of soluble and insoluble fraction separation. Hopp and Wood equation characterized the amino acids on the basis of their charges given positive values to charged residues and negative to aliphatic and aromatic residues (Hopp and Woods 1981).

Table 1.10.17: Example of the soluble and insoluble proteins mathematically evaluated by (Wilkinson and Harrison 1991).

Protein	Resides	Fractions				Net charge	Reference
		Turns	Proline	Cysteine	Hydrophilicity		
CGN4 Ca$^+$ binding domain	105	0.209	0.034	0.000	0.550	0.100	Soluble (Nagai, Thogersen et al. 1988)
Human Tropomyosin	322	0.099	0.000	0.003	0.919	-0.90	
Human α globin	179	0.212	0.039	0.005	-0.014	0.023	Insoluble (Nagai, Thogersen et al. 1988)
Pancreatic Ribonuclease A	169	0.244	0.026	0.049	0.217	0.039	

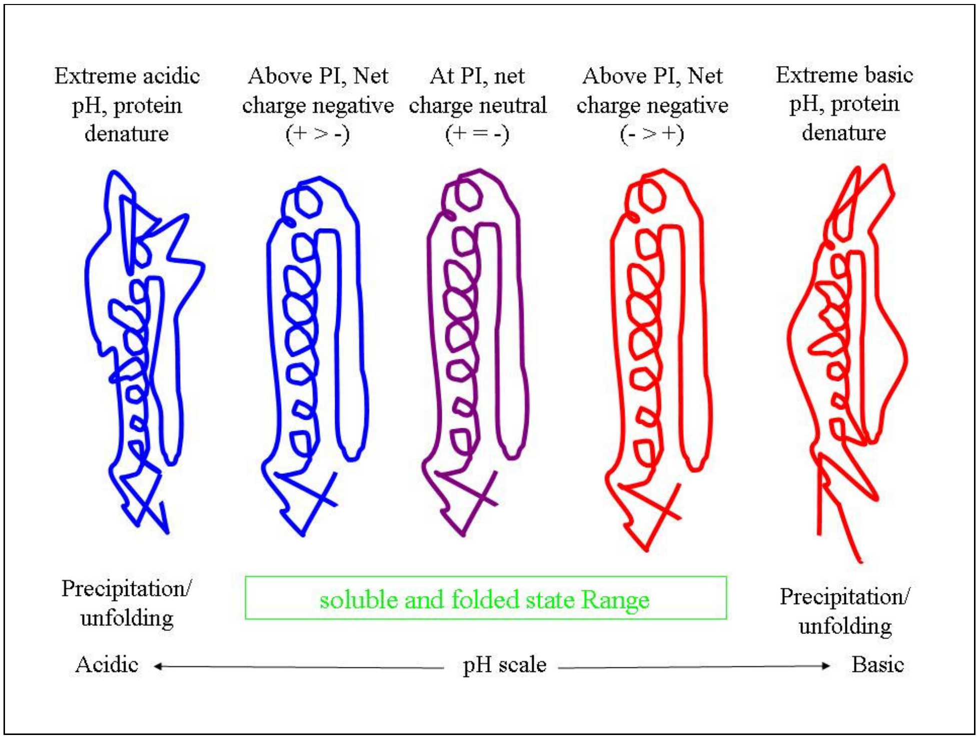

Extreme acidic pH, protein denature
Above PI, Net charge negative (+ > -)
At PI, net charge neutral (+ = -)
Above PI, Net charge negative (- > +)
Extreme basic pH, protein denature
Precipitation/ unfolding
soluble and folded state Range
Precipitation/ unfolding
Acidic
pH scale
Basic

Figure1.10.17.B: Variation in pH under optimal condition for a specific protein can cause the increased net negative/ positive charge, but beyond optimal it can cause protein unfolding/precipitation

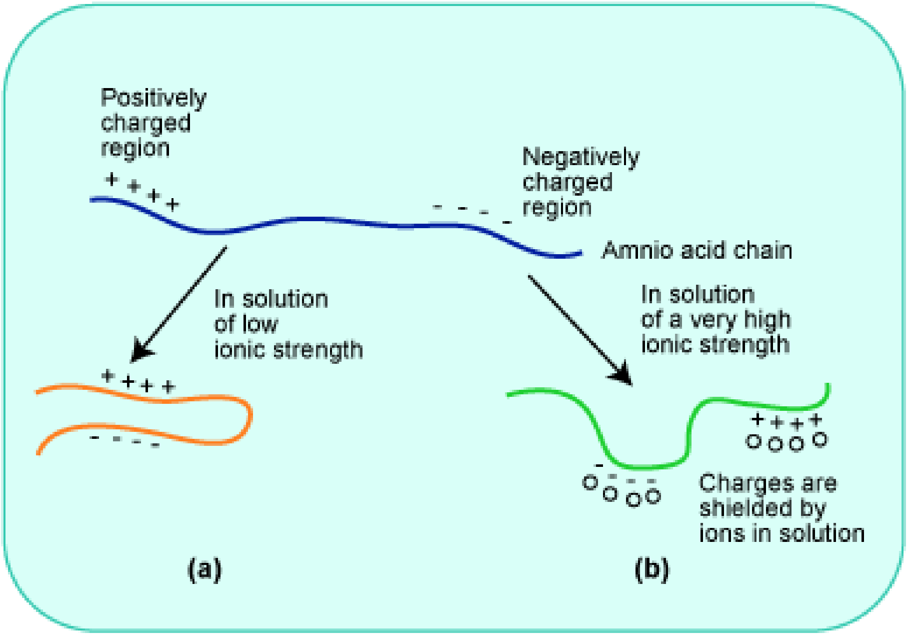

Positively charged region
+ + + +
Negatively charged region
- - - -
Amnio acid chain
In solution of low ionic strength
In solution of a very high ionic strength
+ + + +
- - - -
+ + + +
o o o o
- - - -
o o o o
Charges are shielded by ions in solution
(a)
(b)

Figure 1.10.17.C: Effects of increased ionic concentration on protein solubility and unfolding. (a) Proper salt bridging; (b) salt bridge opened in higher ionic concentration.

Each protein has distinct solubility limitation for the temperature, pH range and salt concentration, depending on their source of origin (environmental adaptation), and amino acid composition, specially the residues that makes the surface of proteins.

Production of proteins with higher hydrophobicity always increases the chances of insolubility. This fact have several proves, membrane proteins usually with higher content of hydrophobic amino acids always suffer with low productivity and insolubility. Luan et al, has done some expression experiments for *C. elegans* ORFeome in high through put fashion. Among 10,167 ORF's, 19 ORFs could not be expressed as they were predicted to have the highest GRAVY (Grand Average Hydrophobicity) values among all, similarly it have been observed from their experimental data that where proteins have more negative GRAVY, yield is not only highest but soluble as well (Luan, Qiu et al. 2004).

Sometime, aggregation cannot be avoided due the presence of hydrophobic residues on the surface of the protein. These residues tends to stick together or with any other substance where they can hide their hydrophobic surfaces away from polar solvents (figure 1.10.17. D). For this kind of proteins, addition of detergents are really helpful to retain their solubility in the buffer. Membrane proteins and proteins that have any hydrophobic surface patch for their special functions, falls in this category.

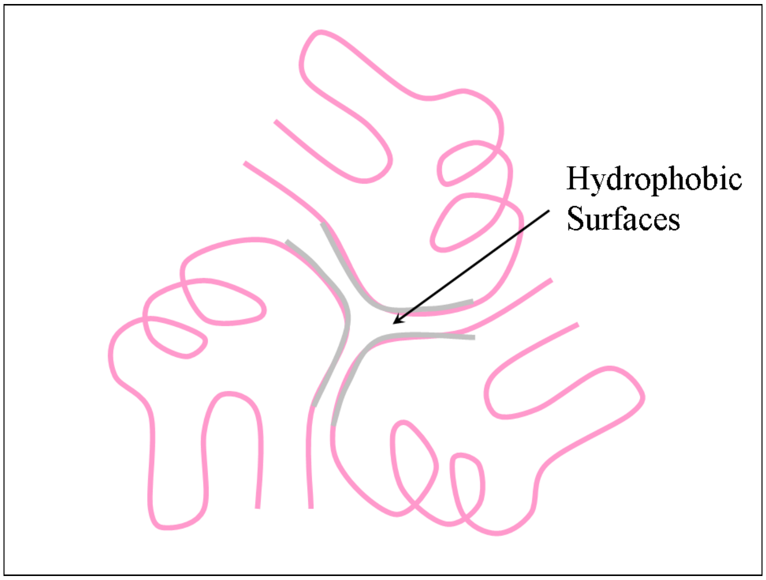


Figure 1.10.1.D: aggregation or clumping due to the hydrophobic surface on proteins.

1.10.18. Possible Solutions to Insolubility:

There could be many solutions to the solubility issue, some time some of the following factors affects greatly to achive the native, active and soluble state of insoluble proteins.

I. Low temperature:

The most common way to take the protein expression toward solubility is to grow cells at lower temperature. There are many example known where lowering the temperature can cause the increased solubility in expressed protein. Such as in case of cloned "k11 RNA polymerase" from *klebsiella* phage expressed in the *E. coli* host, did not turned soluble even in the presence of chaperons until the temperature was decreased to 25°C (Han, Lee et al. 1999). Lowering of temperature can cause the slower production of recombinant protein and in turn less saturation on cell protein folding machinery. But importantly, lowering temperature usually demands more expression time (usually over night) to produce the recombinant protein in significant amount.

II. Reduce expression:

Limited expression is another approach where protein aggregation can be avoided if target protein is express in higher amount and this increased amount in limited space of living cell (mostly in case of *E.coli*) is causing the tangling, jumbled and muddling of target protein (Nuc and Nuc 2006)

III. Solubility tags:

Expression of insoluble target protein in connection with soluble accessory protein (tags) is the most common way to enhance solubility. Connection of soluble protein to the responsible insoluble terminal (hydrophobic a.a containing terminal) can enhance the solubility. The useful solubility enhancing accessory proteins are Z-domain, GST, MBP, Trx, NusA, S- domain, Chloramphenicol Acetyl Transferase and Ubiquitin (Cabrita and Bottomley 2004).

IV. Chaperone mediated solublization:

Co-expression of chaperones to speeding up the folding, in compensation of higher rate synthesis of proteins can assist in soluble expression. Inside the cytosol of *E.coli,* folding of nascent polypeptide is assisted by ribosome-associated Trigger Factors the KJE (DnaK together with DnaJ and GrpE cochaperones) and ESL (GroEL with GroES cochaperone). These two chaperone systems (KJE; ELS) also assist the ClpB (disaggregating chaperone) in solubilization of aggregated proteins. KJE/ClpB also work in close association with small Heat shock proteins (sHSP 31/33) and IbpA/B for disaggregation and refolding of thermally stressed unfolded proteins (de Marco, Deuerling et al. 2007). Figure1.10.18.A explains the chaperones mediated folding intracellularly.

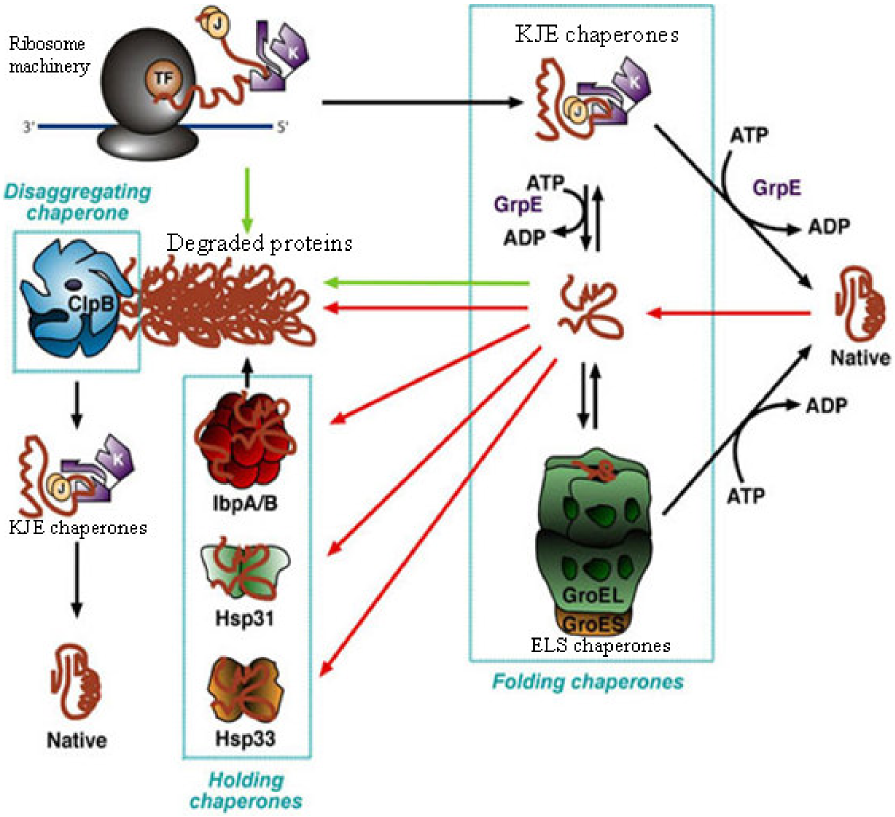


Figure1.10.18.A: Chaperon based folding in cytoplasm of *E. coli*. Red arrow denotes the pathway of thermolabile proteins. Green arrow denotes the aggregation destiny. Folding pathway (black arrow) requires energy in form of ATP, in case of KJE folding chaperone that is provided by GrpE. (Baneyx and Mujacic 2002 ; Baneyx and Mujacic 2004).

Co-expression of chaperones can cause extra burden on cell synthesizing machinery resulting in half or more than half strength of cell gone behind the synthesis of chaperones. But if it is significantly important to assist the solublizations of very insoluble protein then this method can prove beneficial. In recent advancement of chaperone based solubilization assistance, a novel two step method has been described by Marco and coworkers. They have tried a combination of chaperones to be co-expressed for difficult target, solubility enhancement in first step, and then at the end of expression they have attempted to solublize the aggregated remaining by letting chaperones work in combinations of disaggregation chaperone (de Marco, Deuerling et al. 2007).

IV. Protein modification:

Chopping of genetic sequences without any deleterious effects on the protein functionality is another option in case of the well studied proteins where their functions and the functionally active parts are known. This approach sometimes used in high throughput fashions and combinatorial libraries generates, with truncation, point mutation and fragmentation. Based on the phenotypic character of these directed evolutionary clones, selection and rejection take place for elites. (Hart and Tarendeau 2006).

VI. Application of detergents:

When protein surface is unavoidably hydrophobic, then use of detergents can keep the protein stable in solublization buffer. But sometime addition of detergent results in protein inactivation and in most cases inability of proteins to crystallize. Triton X-100, Triton X-114, NP-40, Brij-35, Brij-58, Tween-20, and Tween-80 are Nonionic type of detergents that cannot be dialyzed, while Octyl Glucoside, Octylthio Glucoside (non-ionic type), SDS (anionic type), CHAPS, CHAPSO (Zwitterionic type), can be dialyzed and removed to avoid undesirable properties.

VII. Protein expression in periplasm:

Transport of recombinant protein into periplasm is advantageous and obligatory when protein is toxic or its proper folding needs disulphide bonding, or when down stream purification processes need to be reduced in steps and cost (Makrides 1996). *E.coli* cytoplasm has reducing environment, that do not promote disulphide bonding, therefore nascent polypeptide need to be transported in periplasm where oxidative environment and number of disulfide-binding proteins (DsbA, DsbB, DsbC, and DsbD) and petidyl-prolyl isomerases (SurA, RotA, FklB, and FkpA) promote the appropriate folding of thiol-containing proteins (Shokri, Sanden et al. 2003; Miot and Betton 2004). Signal contaning polypeptide synthesis in cytoplasm there it can be prematurely folded in cytoplasm hance on improper folding destained to be degraded or it can be transported out in periplasm by SecB-dependent pathways (Miot and Betton 2004).

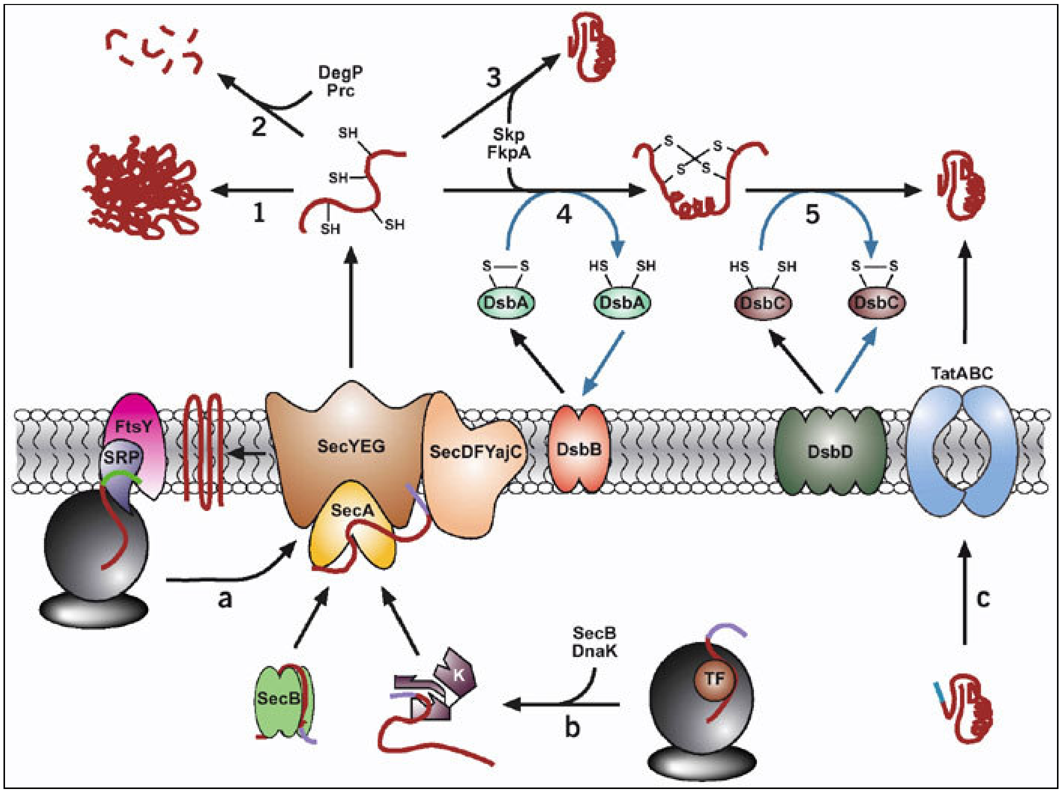


Figure 1.10.2.B: (a) Proteins with highly non-polar sequence (green) recognize by SRP and further transported by SecYEG-SecDFYajC translocons. (b) Proteins with less hydrophobic signal sequence (lavender), co-translational folde by TF and further with assistance of DnaK and SecB transported by Sec-dependent system. (c) Proteins with conserved twin argentine in signal sequence transported by TatABC system, in properly folded form.; (1) The transported protein could have aggregation destiny, (2) degrading fate, (3) proper folding with the help of FkpA, Skp modular, (4) disulphide isomerization by DsbA-DsbB conjugates, (5) DsbC-DsbD based reisomerization in case of incorrect pairing. Source: (Baneyx and Mujacic 2004)

Recombinant protein usually transported in periplasm of *E. coli* by three systems TAT (*t*win-*a*rginine *t*ranslocation pathway), SRP (signal recognition particle Pathway) or SecB-dependent pathways (includes various proteins components SecA, SecB, SecY, SecE, SecG SecD, SecF, YajC, Trigger factor) (Froderberg, Houben et al. 2003; Mergulhao, Summers et al. 2005). Details of these pathways and destinations of signal containing proteins in periplasm have tried to be explained in figure 1.10.18.B.

VIII. Protein transport out of periplasm:

When periplasmic proteins tend to get insoluble in periplasm due to the higher concentration in lesser space then exporting it out of cell is an easier solution for keeping it soluble in an additional advantage of low cost downstream processing during purification (Bieker and Silhavy 1990; Mergulhao, Summers et al. 2005).

After the excretion of target protein in periplasm of *E. coli*, a challenge is to excrete the protein into the media. For this purpose, several attempts have been made ranging from chemical assistance, physical shocks and gentical manipulations. Use of chemicals like glycine and bacteriocin for the enhancement of secretary pathways of *E. coli.* during the expression process have been in use (Yu, Aristidou et al. 1991). Other conditions like modrate copy number of plasmid and use of minimal media M9 is also reported to enhance the soluble secreted protein (Mergulhao, Monteiro et al. 2003). Abrahamsen et, al. has reported that the insertion of heat shock response protein prior to the gene of interest can enhance the secreted expression (Abrahmsen, Moks et al. 1986). In genetical approach, a novel pComb3 phagemid system have been discovered, that aims to produce antigen and antibodies in mg/l quantities on the surface of phagmoid transfected to *E. coli* (Barbas, Kang et al. 1991).

Five different types of secretion systems work in gram negative bacteria, among them only Sec-dependent pathways contains cleavable signal peptides. In non Sec-dependent pathway only HlyA (hemolysin A type secretion) directly transfer the *C*-terminally attached protein into the media with simultinous disulphide bonding (Kostakioti, Newman et al. 2005).

IX. Re-solublization of inclusion bodies:

Some scientists prefer to express the protein as inclusion bodies because they are easier to obtain with higher purity (~ 90-95%) just with a one step of centrifugation. In case of inclusion bodies formation, very high level of expression (>30% of cell weight) is possible which favorably remains prevented with the cell proteolytic degrading machinery (figure: 1.10.18.C). A disadvantage of expressing proteins in inclusion bodies is the absence of active protein, so real time monitoring of expression by activity assays is not possible (Das and Mukhopadhyay 1994). Denaturing and renaturing protocol can be applied on inclusion bodies with no surety of success for all target proteins. Often these attempts result in higher cost and labor intensive initial experiments (Cabrita and Bottomley 2004).

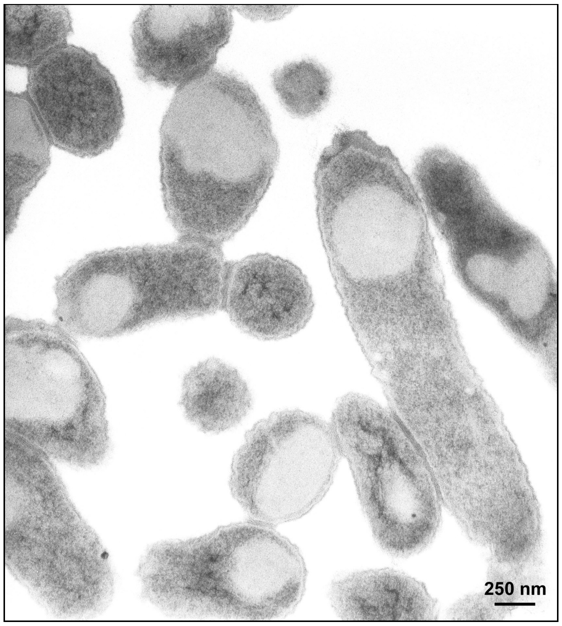


Figure1.10.18.C: electron microscopic photograph of inclusion bodies formation in *E.coli* cells. Source: www.boku.ac.at

X. Composition of lysis buffer:

Some proteins or protein fractions fall into the insoluble fraction due to the inappropriate buffer conditions, even if they are not in inclusion bodies. Therefore a trial must be run for every new protein to look for appropriate salt, pH, and stabilizers conditions.

Aims & Objectives

One of the aim of this project was to perform genome wise bioinformatic analysis of "Aliivibrio salmonicid" to search for degradome related ORFs. This information includes all proteases, protease inhibitors and proteins that could have protease activity beside their other functions. The purpose of these bioinformatic studies was to find out the unique sequences with lesser similarity to any known PDB structures. Looking for the less similar sequence from the genome pool is an important tool to detect or discover the new/novel folds. Aiming toward these targets could be interesting to reveal new/novel folds through X-ray crystallography.

Beside the novelty of structure, it was also a secondary aim to obtain some proteases with unique specificity or could have uniqness in prerequisite catalytic environments. These aims further linked with the commercial aspects of chosen targets that arise from cold adapted habitats. It was expected that enzymes purified from cold adapted habitants will have some qualities able to exploit commercially (D'Amico, Collins et al. 2006).

Cloning and protein expression was the main tool for producing chosen targets for functional and structural studies. The problems encountered during the expression and purification, of these targets proteins need to be optimized for the economically feasible protein production. These encountered problems include more general insolubility and more specifically cold habitant associated hydrophobicity (Thorvaldsen, Hjerde et al. 2007).

Beside these main objectives of identifying proteases from the genome of *Aliivibrio salmonicida*, functional prediction was another motivation. Blast results into the non redundant NCBI data-base was an important tool for this purpose. This functional prediction can be ultimately correlate with the protease based virulence factors (Goguen, Hoe et al. 1995; Araiza Orozco, Avila Muro et al. 1997; Lantz 1997).

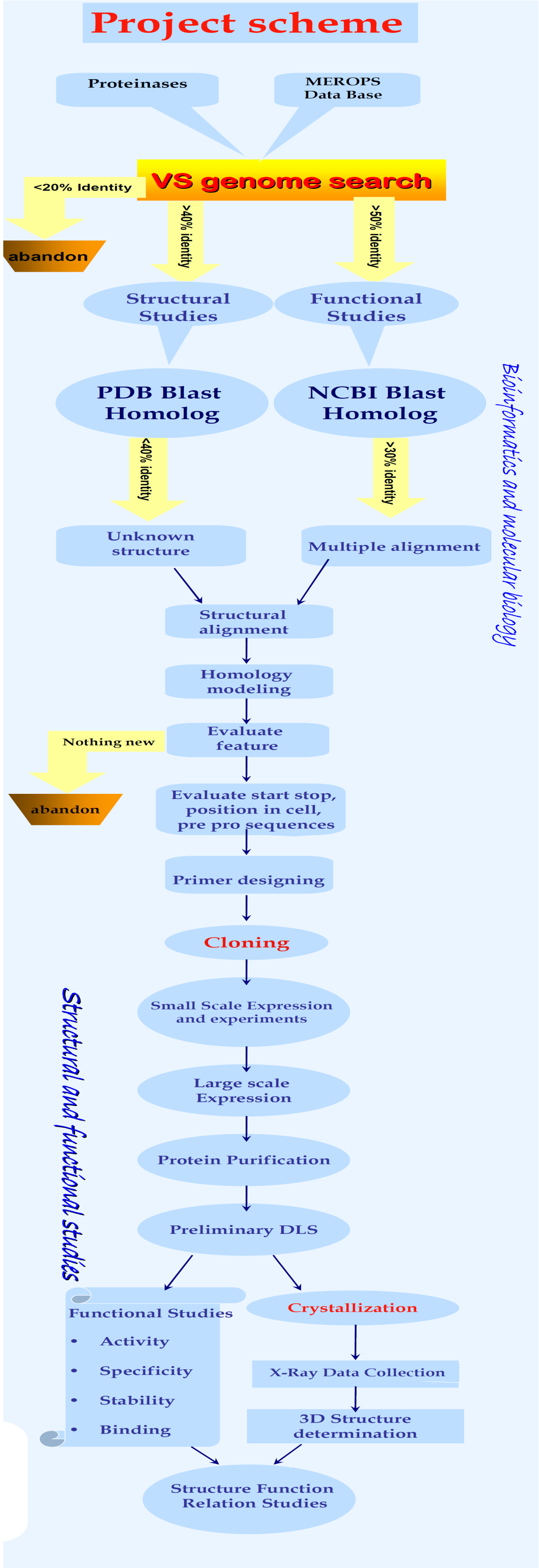


Figure 2.1: schematic representation of aims and objectives.

Experimental

3.1.0. Bioinformatical analysis of *V.salmonicida* genome

This project was an integrated approach of multidisplanary technologies. Initiated with the bioinformatics approach to analyze the *Vibrio salmonicida* genome for the presence of degradomic contents. The genome sequence from *A. salmonicida* was converted into the ORF through EBI server (*http://www.ebi.ac.uk/Tools/emboss/transeq/index.html*) and then further blast into the MEROPS database (*http://merops.sanger.ac.uk/*), to identify the possible protease sequences.

171 ORFs were predicted to contain proteolytic activity and 74 sequences possesed protease inhibition activity. The predicted sequences were further analyzed by their functional (NCBI nr Data Base) and structural homology (PDB Data Base) into the NCBI server (*http://blast.ncbi.nlm.nih.gov/Blast.cgi*). Motifs were also identified through Conserve Domain Data Base (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml). Obtained hits were then again abridged on the basis of significance in E-values. 113 proteins found to contain protease activity as a major function or as a part of their structure and function in clear indication which also included the sequences that were similar to unassigned peptidases.

Targets were selected on the basis of:

1) Uniqueness in fold.

2) Possible uniqueness in specificity.

3) Possible uniqueness in activity conditions.

4) Possible involvement in pathogenicity.

5) Commercial potential.

Based on the above criteria, 8 different targets were selected initially. Interesting proteases and protease containing proteins selected to be worked on, included LexA, RadA, ThiJ, HslV, Trypsin, Collagnase, ProtinaseK and ToxR homolog from *A. salmonicida.* Initially selected targets were analyzed thoroughly with structural alignment and homology modeling. At that point prediction server like

Protparam tool (http://ca.expasy.org/tools/protparam.html) was utilized to predict important features of targets, SignalP (http://www.cbs.dtu.dk/services/SignalP/) was utilized to predict the signal sequences, THMMH (http://www.cbs.dtu.dk/services/TMHMM/) was utilized to check the possible transmembrane helices.

Due to the commercial significance of trypsins, TVS4041 was further selected as the target that needs to work in more extensive way.

3.2.0. Gateway™ (Invitrogen) Cloning:

Three different constructs were decided to be made for looking into the variation in expression level, changes in solubility and for the facilitation of purification. They were: (1) Native Construct, to produce the protein in its natural form; (2) *N*-terminal construct, to produce the protein with *N*-terminal 6xHis attachment; (3) *C*-terminal Construct, to produce the protein with *C*- terminal 6xHis attachment. Primer were designing to flank the ORF with attB recombination sites (attB1 and attB2) that was required for specialized cloning technique of the Gateway technology. For each target, four different primers were designed for three different construct preparations. General primer design was like follows:

Gene specific 12*att*B1: 5' AA AAA GCA GGC TNN (Gene specific sequence) 3'
Gene specific 12*att*B2: 5' A GAA AGC TGG GTN (Gene specific sequence) 3'
Universal *att*B1 adapter primer: 5' G GGG ACA AGT TTG TAC AAA AAA GCA GGC 3'
Universal *att*B2 adapter primer: 5' GGG GAC CAC TTT GTA CAA GAA AGC TGG GT3'

Table3.2.0: Gene specific primer for selected targets.

Where vector related sequences (blue), gene related sequences (pink).

Target	Forward native primer	Reveres native primer	N-terminal His tag primer	C-terminal His tag primer
LexA	AA AAA GCA GGC TTC GAA GGA GAT AGA ACC **ATG** AAG CCA CTA ACG GCA AGA CAG CAA G	A GAA AGC TGG GTC **TTA** CAT CCA AGT CGT ACT ACG GAT CAC	AA AAA GCA GGC TTC **ATG** AAG CCA CTA ACG GCA AGA CAG CAA G	CAA GAA AGC TGG GTC CAT CCA AGT CGT ACT ACG GAT CAC
RadA	AA AAA GCA GGC TTC GAA GGA GAT AGA ACC **ATG** GCA AAA GCA AAA CGT GCC TAC GTT TGT AAC GAT TG	A GAA AGC TGG GTC **TTA** TAG TTC ATC AAA AGC ATC AAT TG	AA AAA GCA GGC TTC **ATG** GCA AAA GCA AAA CGT GCC TAC GTT TGT AAC GAT TG	CAA GAA AGC TGG GTC TAG TTC ATC AAA AGC ATC AAT TG
ThiJ/ PfpI	AA AAA GCA GGC TTC GAA GGA GAT AGA ACC **ATG** GAA ATG AAA AAA ATT GCG	A GAA AGC TGG GTC **TCA** GCT TTC GGA TAT CAC TTT TTC	AA AAA GCA GGC TTC **ATG** GAA ATG AAA AAA ATT GCG	CAA GAA AGC TGG GTC GCT TTC GGA TAT CAC TTT TTC
HslV	AA AAA GCA GGC TTC	A GAA AGC TGG	AA AAA GCA GGC	CAA GAA AGC

	GAA GGA GAT AGA ACC **ATG** GAG GTT TTA CTC GTG ACT AC	GTC C **CTA** TTT GTT GGT ATC AAG AAC TTC	TTC **ATG** GAG GTT TTA CTC GTG ACT AC	TGG GTC TTT GTT GGT ATC AAG AAC TTC
Trypsin	AA AAA GCA GGC TTC GAA GGA GAT AGA ACC **ATG** AAC GTT GTT GTA GGA GCA CTT GTC TCG	A GAA AGC TGG GTC **TCA** TGA CGC TCT ACT CTT TCG GTA G	AA AAA GCA GGC TTC **ATG** AAC GTT GTT GTA GGA GCA CTT GTC TCG	A GAA AGC TGG GTC TGA CGC TCT ACT CTT TCG GTA G
Collagnase	AA AAA GCA GGC TTC GAA GGA GAT AGA ACC **ATG** ATA ATT AAC AAT TAT ATC AGC AGT TGG	GAA AGC TGG GTC **CTA** TTC GTA TGT GAC GTA TAC TTC CG	AA AAA GCA GGC TTC **ATG** ATA ATT AAC AAT TAT ATC AGC AGT TGG	GAA AGC TGG GTC TTC GTA TGT GAC GTA TAC TTC CG
Protinase K	AA AAA GCA GGC TTC GAA GGA GAT AGA ACC **ATG** AAA AAT ATT AAA CGC TCA TTC ATG TGC	A GAA AGC TGG GTC **TTA** CCA TTC TTC TAA TGA TGG TTT AGG	AA AAA GCA GGC TTC **ATG** AAA AAT ATT AAA CGC TCA TTC ATG TGC	A GAA AGC TGG GTC CCA TTC TTC TAA TGA TGG TTT AGG
ToxR	AA AAA GCA GGC TTC GAA GGA GAT AGA ACC **ATG** CAA CAA CAA CAA TTA TCC	A GAA AGC TGG GTC **TTA** TTG GGT AAG TTG TTG ATT ATT TTT TTG	AA AAA GCA GGC TTC **ATG** CAA CAA CAA CAA TTA TCC	A GAA AGC TGG GTC TTG GGT AAG TTG TTG ATT ATT TTT TTG

3.2.1. Two Steps Gateway PCR for ORF flanking with attB1 and attB2 sites:

Gene amplification: DNA for each construct was prepared from the isolated genome of *A. salmonicida* with the application of high fedility Platinum® *Pfx* DNA polymerase proof reading enzyme (**invitrogen**). Step one brought a small part of *att* site through 12*att*B gene specific primers, which was extended in step two with the universal 12*att*B adapter primers.

Step one standard PCR reaction contained 5µl Pfx buffer (10x), 1.5µl dNTPs Mix (10mM), 1.0µl 50mM $MgSO_4$, 1.5µl 10µM 12attB1 primer, 1.5µl 10µM 12attB2 primer, 1.0µl 10pg-200pg genomic DNA, 1.0µl 1.25units Platinum *pfx* Polymerase in 50µl reaction. Reaction was performed in thermocycler (Perkin Elmer). Step one standard conditions were 95°C for 5 mins, followed by 25 cycles of 5sec at 94°C (denaturation), 15sec at 55°C (annealing), 2min at 68°C (extension), before ending an extended step at 68°C for 7 min and ended up with 4°C cooling.

Evaluation of PCR products were done through agarose gel Electrophoresis. PCR products were run on 0.8% agrose gel (0.8g/ 100 ml TBA buffer, 4µl ethidium bromide /100ml) and checked for the correct weight PCR product. In case of multiple, products were isolated from gel in correct band size and purified through QIAEX II Agarose gel extraction Kit. In case of single product, gene product was purified through QIAGEN QiaQuick PCR purification kit.

Step two standard PCR reaction contains 5µl Pfx buffer (10x), 1.5µl dNTPs Mix 10mM, 1.0µl $MgSO_4$ (50mM), 1.0µl attB1 primer (10µM), 1.0µl attB2 primer (10µM), 10.0µl Purified PCR product from step one, 1.0µl Platinum *pfx* Polymerase (1.25units) to a volume of 50µl. Reaction was performed in thermocycler (Perkin Elmer). Step two standard conditions were in two phases, phase one had 20 cycles with 30sec at 95°C (denaturation), 30sec at 55°C (annealing), 1min at 68°C (extension), followed by extending step at 68°C for 5 min and ended up with 4°C cooling.

Phase two of step two was started with initial denaturation for 5 min at 94°C and then 5 cycles with 30sec at 94°C (denaturation), 30sec at 55°C (annealing), 1min at 68°C (extension), followed by an extended step at 68°C for 5 min and ended up with 4°C cooling. PCR products were again evaluated for quality on agarose gel and purified as describe above.

3.2.2. BP recombination reaction:

BP reactions were performed to bring the genes of intrest in Lysogenic pathway. The reaction was done as describe in the Invitrogen® Gateway's protocol, summerize as below:

*att*B1-Target Gene- *att*B2 + *att*P1-ccdB-*att*P2 ➔ *att*L1- Target's Gene -*att*L2 +*att*R1-ccdB-*att*R2 *(attB flanked ORF)* + *(Entry vector)* ➔ *(Entry clone)* + *(by-Product)*

8µl of TE Buffer pH 8.0, 1µl of empty entry vector pDONOR221 (150ng/µl), 2µl BP clonase™ II Enzyme Mix [bacteriophage λ Integrase (Int) and *E. coli* integration host factor (IHF)] and 10ng/µl of *att*B-PCR product was briefly vortexed mix and incubated at 25°C for 60 min or overnight for completion of reaction. Reaction was stopped by adding 1µl protinaseK for 10 min at 37°C.

1-5µl of reaction mixture was transformed into the 50µl of aliquot of subcloning efficiency DH5α competent cells. Transformed entry clone were selected on LB plates provided with Kanamycin 50µg/ml. From transformed Kanamycin^R plate, colonies were selected for plasmid isolation.

An overnight grown culture was harvested for plasmid isolation with "QIAGEN Plasmid Isolation Kit" and observed on gel for the presence of plasmid. Colonies containing plasmid were marked and preserved for further use. From the isolated plasmid a PCR was run to verify the presence of target gene in the plasmid and observed on gel with their relevant base pair sizes. Hence the plasmid containing genes were conserved for LR reaction. Prior to LR reaction a sequencing step was done to confirm the correct frame of inserted gene in the vector's translational frame. Sequencing reactions were done using v3.1 Big Dye Chemistry™ Kit (Applied Biosystem). Reaction mixture contained 150-300ng plasmid DNA, 2µl Fwd/ Rev Primer (4mM), 2µl sequencing Buffer, 2µl Big Dye v3.1 enzymes and milliQ water to 20µl total.

<u>Sequencing PCR cycling parameters</u>

step1.	96°C	-1min
step2.	96°C	-10Sec
step3.	50°C	-5Sec
step4.	60°C	-1min
Step 2-4	34 times more	
step5.	60°C	-7min
step6.	4°C	-incubation

3.2.3. LR Recombination reaction:

LR reaction was performed to bring the target gene into the lytic pathways. The reaction was done as described in the Invitrogen® Gateway's protocol, summarize below:

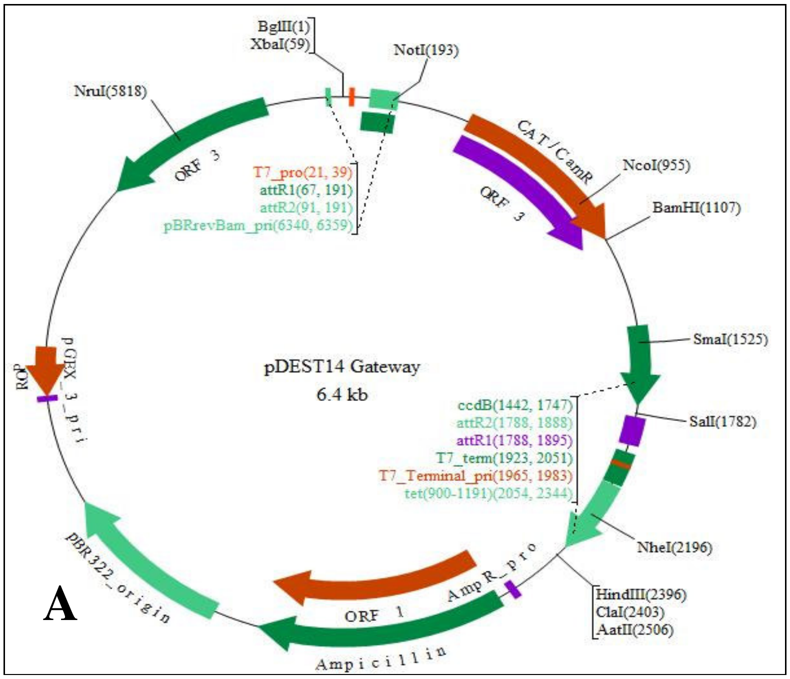
*att*L1- Target's Gene -*att*L2 + *att*R1-*ccd*B-*att*R2 ➔ *att*B1- Target's Gene - *att*B2 + *att*P1-*ccd*B-*att*P2

(Entry clone) + *(Destination vector)* ➔ *(Expression clone)* + *(by product)*

Reaction was catalized by 1µl empty destination vector (150ng/µl), 4µl LR Clonase™ II Enzyme Mix [bacteriophage λ Integrase (Int), Excisionase (Xis) and E.coli integration host factor (IHF)], 50-150ng/µl Entry clone and 8µl TE buffer pH 8.0. While, empty destination vectors used were: pDEST14 for native construct / pDEST17 for *N*-ter His tag construct pET DEST42 for *C*-terminal His tag construct. Details of the vectors are given figures 3.2.3 (A, B and C) respectivly (Source: www.biovisualtech.com).

Reaction mixture was mixed and incubated at 25°C for 60 min or over night. Reaction was stopped by mixing 1µl protinase K at 37°C for 10 min. 1-5µl of reaction mixture was transformed into the 50µl of aliquot of subcloning efficiency DH5α (as non expression, multiplication host) and BL21 (DE3) CodonPlus-RIL/ BL21 AI competent cells (as expression hosts).

Ampicillin^R Transformed entry clone were selected on LB plates provided with Ampicillin 100µg/ml. From an over night grown plated transformed LR clones were selected and inoculated in a broth for plasmid isolation. Gene detection and confirmation were done as describe above in case of BP reaction.



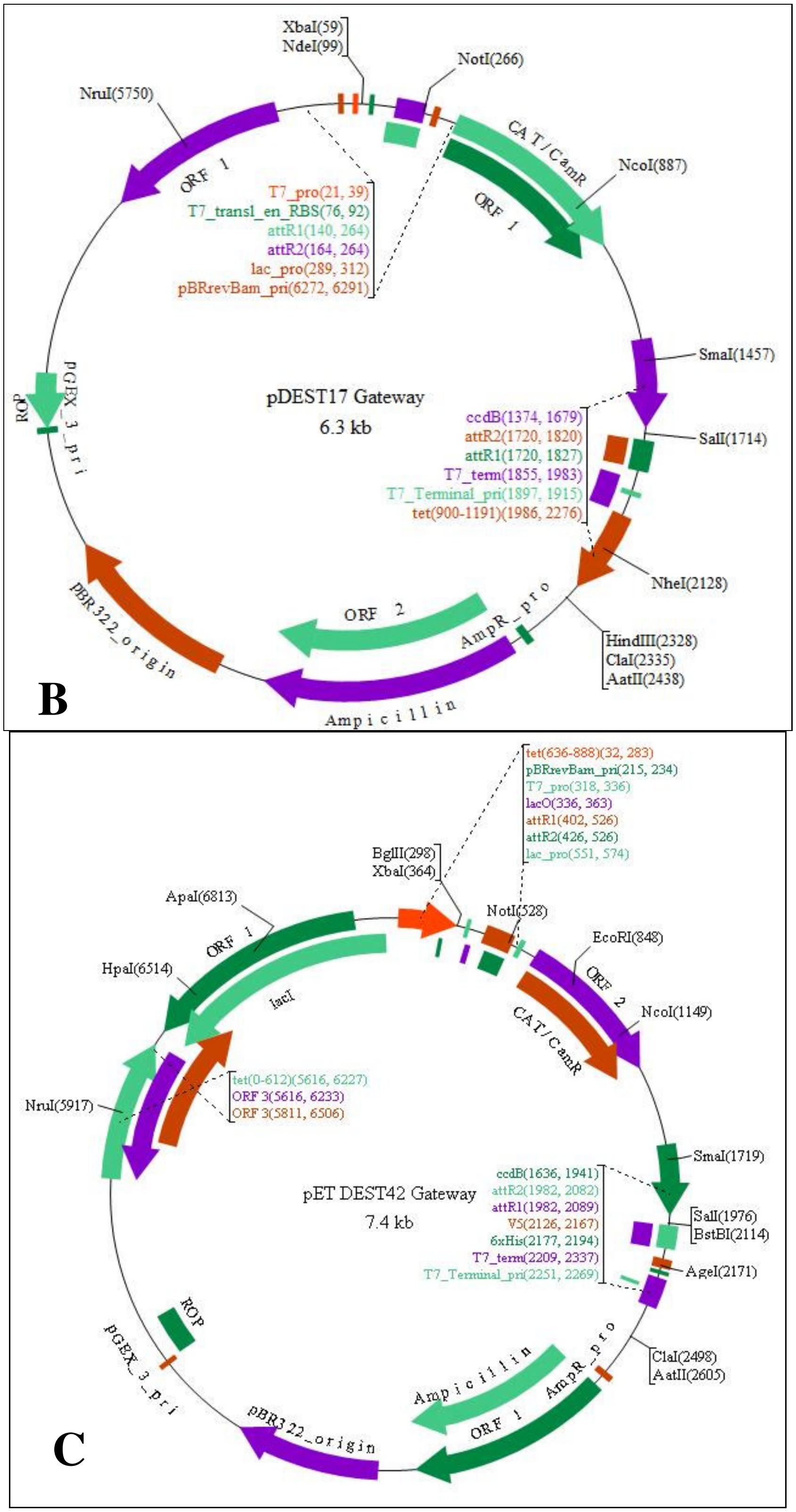


Figure 3.2.3: Expression plasmids Map, pDEST 14(A), pDEST 17 (B), pET DEST 42 (C);
(Source: www.biovisualtech.com)

3.3.0. Restriction digestion cloning (pBADgIII/TOP10 cloning):

For TVS4041 an additional cloning method was adapted, which leaded target protein to periplasmic region with the assistance of a vector signal sequence.

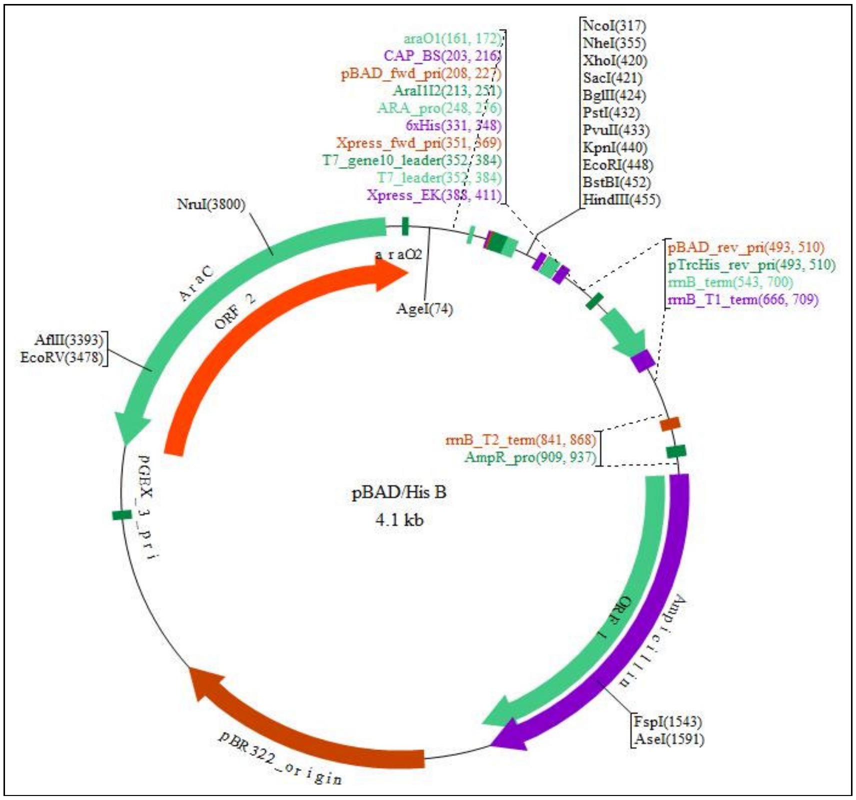


Figure 3.3.0: Expression plasmids Map, pBAD/gIII(B), invitrogen
(Source: www.biovisualtech.com)

TVS4041 (Trypsin gene sequence) was analyzed through software (http://tools.neb.com/NEBcutter2/index.php) for possible restriction digestion sites. Restriction sites ***NcoI*** and ***SalI*** were not present in the gene sequence and they were on extreme *N* and *C*-terminals of vectors' translational sites, therfore by choosing them as cloning site, minimal extra codones could be translated with the protein.

3.3.1. Primer designing and PCR:

Primers were designed with ***NcoI*** and ***SalI*** sites in *N*- and *C*-terminal respectively. The *N*-terminal was selected on the target gene without Signal sequence (Pre-sequence) from *A. salmonicidal* source; two different origins were selected for initiation site, a 10 a.a Pro-sequence contaning site (for zymogen encoding) and without Pro-sequence contaning (maturate enzyme). They were named as 1 and 2 respectivly. The plan was to make 4 construct with pro-sequence native (N1), without pro-sequence native (N2), pro-sequence His tagged (H1) and without pro sequence His tagged (H2).

Following primers were designed, where ***NcoI*** and ***SalI*** sites are represented in red color, additional bases necessary to bring the cloned gene in frame represents in green color.

N-terminal 1: GC A*CC ATG G*CA ACC GAA GAG TTT TCA GTT ACT CC
N-terminal 2: GC A*CC ATG G*GC ATT GTG GGT GGT AAT GAC GCG
C-terminal Native: AAT *GTC GAC* **TCA** TGA CGC TCT ACT CTT TCG G
C-terminal 6xHis Tag: AAT *GTC GAC* TGA CGC TCT ACT CTT TCG G

Gene specific amplication was done with the primer describe above. Through this PCR ***NcoI*** and ***SalI*** sites were added in the *N*- and *C*- terminal of gene to be cloned. PCR reaction mixture was prepared with 5µl Pfx buffer 10x, 1.5µl dNTPs Mix 10mM, 1.0µl $MgSO_4$ 50mM, 1.5µl *N*-terminal primer 1 or 2 10µM, 1.5µl *C*-terminal Native/ 6xHis Tag primer 10µM, 1.0 µl 10-200 pg genomic DNA, 1.0µl Platinum *pfx* Polymerase 1.25units in a 50µl reaction.

3.3.2. Restriction digestion & ligation:

All four amplified constructs and pBAD/gIII vector was double digested with the ***NcoI*** and ***SalI*** sites in five separate tubes. Digestion mixture contained 18µl of DNA different constructs or 2µl (pBAD/gIII vector), 3µl of NEB buffer3, 3µl of 1mg/ml BSA, 2µl of ***NcoI*** (10.000 U/ml), 2µl of ***SalI*** (20.000 U/ml) and 2µl distilled Water. Reaction mixture was incubated at 37°C for 1hr and then deactivated at 65°C for 10min. digested products were purified with the PCR cleanup kit QIAquick™. For all five digested products DNA concentration was measured through Nanodrop® machine.

Required amount of gene to be ligated with the 10ng of vector was calculated 8.897 ng for pro-sequence construct and 8.634 ng for without pro-sequence construct, by using following formula:

$$X\text{ng insert} = \frac{4\ (\text{bp insert})\ (10\text{ng linerized vector})}{\text{bp vector }(4100)}$$

Ligation mixture contained calcutated amount of ***NcoI*** and ***SalI*** digested gene products, 10µl of ***NcoI*** and ***SalI*** digested pBAD/gIII vector, 1µl of ***T4 DNA ligase*** Buffer, 1µl of ***T4 DNA ligase*** and milli Q water to 20µl total volume. Reaction mixture was incubated for ligation in PCR at 16°C for overnight.

3.3.3. Transformation and clone selection:

For transformation of ligated product in one shot chemically competent TOPO10™ (invitrogen) cells, 1-5µl of ligated product was gently mix with brief thawed cells on ice. After 30 minutes incubation on ice, cells were transformed by heat shock at 42°C water bath for exactly 30 seconds. Cells were kept on ice for 2 mins and then added 250 µl of pre-warmed S.O.C medium. Cells were grown in 225 rpm shaking incubator at 37°C for exactly 1 hour. Transformed were selected on LB agar plates containing ampiciline 100µg/ml, after a night incubation.

A PCR Master Mix (AB gene™) in addition with gene specific primers was used to perform the colony PCR. Colonies were marked and picked for direct utilization in the PCR tubes. Plasmids were exposed by boiling in the initial stage of PCR. Following PCR conditions were used to check the correct size Genes inserts into the vector.

step1. 96°C -5min (initial denaturation)
step2. 96°C -30Sec (denaturation)
step3. 50°C -30Sec (Primer annealing)
step4. 65°C -1min (Extension)
Step 2-4 34 more time
step5. 65°C -5min (Final Extension)
step6. 4°C -incubation

Correct size of inserts was checked on 0.8% Agarose gel run.

Clones were sequenced to check the correct frame and any mutation in the gene sequence. Sequencing reactions were done using v3.1 Big Dye Chemistry™ Kit (Applied Biosystem). Reaction mixture contained 150-300ng Plasmid DNA, 2µl 4mM Fwd/ Rev Primer, 2µl sequencing Buffer, 2µl Big Dye v3.1 enzymes, MQ water to 20µl total. Sequencing PCR cycling parameters were the same as describe in section 3.2.2.

3.4.0. Solubility improvemnt of TVS-4041:

TVS4041, tried to express with pBAD/gIII vector was optimized in different cultural conditions to see the possible enhancement of solubility, that was not very good in general expression conditions.

For all of the expression studies, an overnight culture was grown to OD_{600} ~ 1-2, and 1:10 volume was inoculated usually in 10ml – 50ml fresh medium containing ampicillin 100 µg/L. For all the treatments, OD_{600} was measured on harvest to check the toxic effects and to bring all the experimental treatments in a same standard (La Vallie 1993). Formula for determining volume for resuspension of cell pellet is given as under.

$$\mathbf{V_{resuspension} = (OD_{600}\ of\ culture/\ 5)\ x\ volume\ of\ culture}$$

All of these results were analyzed on Western blot and additionally to SDS-PAGE. Activity measurements were done using BAPNA (sigma Aldrich) as a standard substrate for trypsin activity.

3.4.1. Optimal cell growth stage for induction:

To determine the effects of time of induction on yield and solubility of protein, an experiment was design to induce the cell on different growth stages of the bacterium, indicated with varying OD level. These induction was performed at OD_{600} ~0.6, 0.8, 1.0 1.2 and 1.3.

3.4.2. Inducer concentration optimization:

Ten different concentrations of L-arabinose were tested for optimization of inducer concentration. 10ml of overnight culture was use to make 100ml of culture, grown at 37°C up to the OD_{600}~1.0 then it was divided into 9 culture flask containing 10 ml culture and labeled as 0%, 0.1%, 0.2%, 0.3%, 0.5%, 0.7%, 0.9%, 1.2%, 1.4% Arabinose. These cultures were then induced with different volumes of 20% arabinose to make the final concentration describe above.

3.4.3. Time scale parameter evaluation for maximal protein production

Time scale evaluation for the optimal protein acquirement is essential in all protein expression studies. Based on that fact an experiment was designed and run according to time scale. An overnight culture was used in dilution of 0.1 to make a 50 ml broth at OD_{600} ~ 1.0.

Then it was induced with final volume of 0.2% arabinose and harvest at 0 hr, 1hr, 2hr, 4hr, 5hr, 6hr and 22hr after induction. All these treatments were processed for solubility determination and freezed at -20°C, until analyzed.

3.4.4. Effects of varying NaCl concentration:

NaCl is known to bring effect on solubility of proteins and since the target protein is from halophilic bacterium therefore it is quite essential to evaluate the effects of different salt concentration in medium. These concentrations were 0.1%, 0.5%, 1.0%, 1.5%, 2.0%, 2.5%, 3.0%.

3.4.5. Effect of pH on protein quantity and quality

pH can bring drastic effect not only on normal cell growth but on the protein folding and secretion in to the medium. A study was planned to work out with the varying pH of medium in the normal range to evaluate the possible effect on target protein and factors relating to its quality (San, Bennett et al. 1994).

3.4.6. Temperature optimization:

Temperature can bring drastic effect on metabolism rate in an organism and quantity of target protein and its proper folding rate. This effect was studied with five different temperature treatments in the range between minimum to highest known survival temperature. Cultures were grown at 25°C up to the OD_{600}~1.0 and induce with the L-Arabinose at the final concentration of 0.2%. Five different temperature levels were used to find the optimal temperature to obtain most soluble protein. These temperatures were 16°C, 23°C, 30°C, 37°C and 42°C.

3.4.7. Effect of rear metals in medium:

Metals have known effects on stability of proteins, especially in case of proteases. Some proteases required metal ions for their optimal activity. Therefore Ca^{++}, Zn^{++}, Mn^{++}, Mg^{++}, K^{+} and metals mix (1000x formulation given in table 3.4.7) were added in medium to evaluate the effects on protein expression and yield.

Table 3.4.7: 1000x metal Mix formulation

components	volume
water	36 ml
0.1M $FeCl_3.6H_2O$	50 ml
1M $CaCl_2$	2 ml
1M $MnCl_2.4H_2O$	1ml
1M $ZnSO_4.7H_2O$	1ml
0.2M $CoCl_2.6H_2O$	1ml
0.1M $CuCl_2.2H_2O$	2ml
0.2M $NiCl_2.6H_2O$	1ml
0.1M $Na_2MoO_4.2H_2O$	2ml
0.1M $Na_2SeO_3.5H_2O$	2ml
0.1M H_3BO_4	2ml

3.4.8. Other experiments:

In getting the answer of some questions arises during above experiment, following experiment were also evaluate: Insolubility screening with increasing time of expression, secretion of target protein into the media, effects of glucose for production of target protein, ampicillin and carbencillin for long period induction, effects of aeration on solubility enhancement, comparison of all constructs' soluble expression in modified conditions.

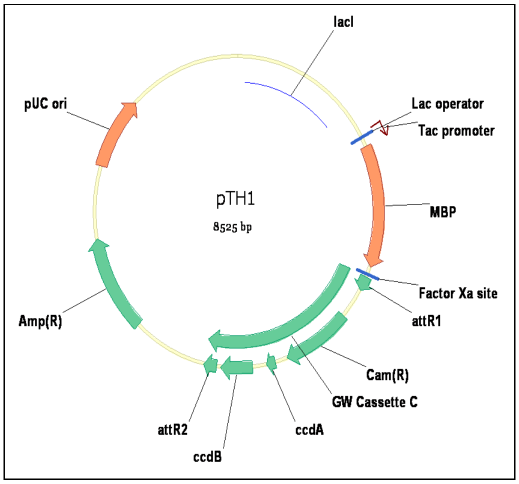
3.5.0. Expression studies of LexA:

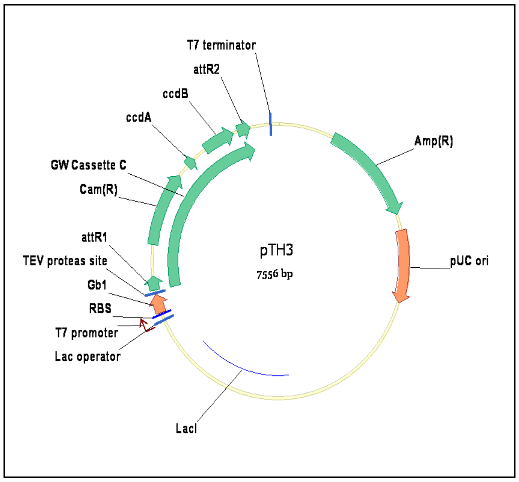
The target protein LexA was tried to optimize in yield by providing different fusion protein in *N*-terminus. For this purpose *N*-terminal construct was cloned into eight different expression vectors. Among them pDEST-TH1, pDEST-TH3, pDEST-TH6, pDEST-TH7 and pDEST-TH10 were gateway adapted *N*-terminal fusion tag vectors, kindly gifted by Dr. Helena Berglund to the Protein Research Group, Biotechnology department university of Tromsø.

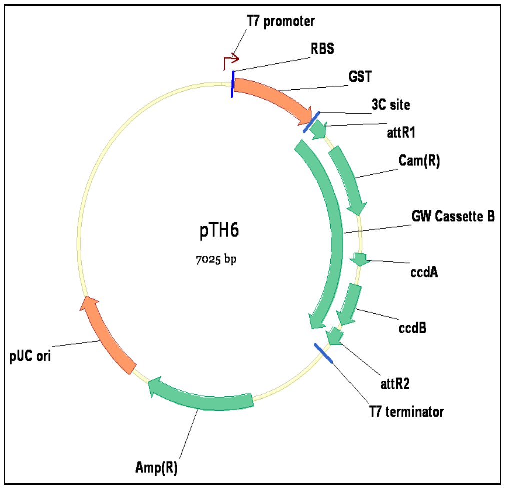
The other three vectors pDEST15, pDEST16 and pDEST17 were from invitrogen. Detail of these expression vectors and their fusion tags are described in table as follows. Further detail can be seen through the vector maps given in figure 3.5.0.

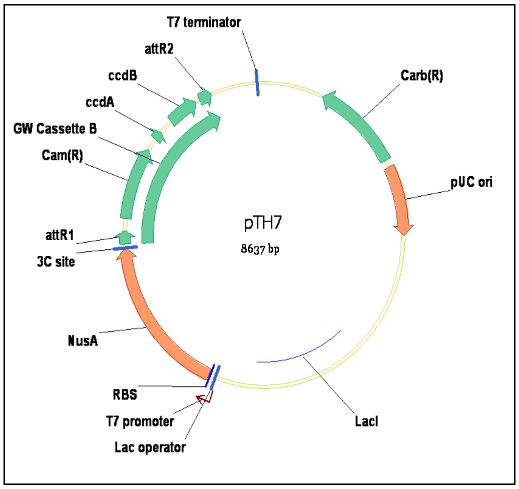
Table 3.5.0: Gateway adapted vectors used for different N-terminal fusion partner in LexA expression.

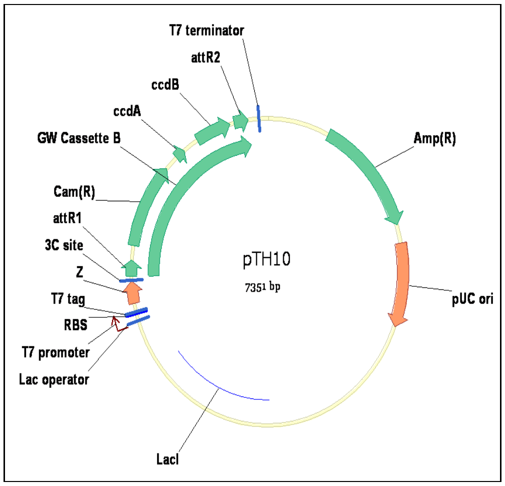
Expression Vector	N-terminal tag	origin	Weight (KD)	Promoter	Cleavage site	Selection marker	Reference
pDEST-TH1	MBP	pUC	44.2	tac	Factor Xa	Amp^R	(Hammarstrom, Hellgren et al. 2002)
pDEST-TH3	Gb1	pUC	8.6	T7lac	TEV	Amp^R	(Hammarstrom, Hellgren et al. 2002)
pDEST-TH6	GST	pUC	~ 26	T7	PreScission(3C)	Amp^R	(Hammarstrom, Hellgren et al. 2002)
pDEST-TH7	NusA	pUC	56.4	T7lac	PreScission(3C)	Cab^R	(Hammarstrom, Hellgren et al. 2002)
pDEST-TH10	Z- domain	pUC	17	T7lac	PreScission(3C)	Amp^R	(Hammarstrom, Hellgren et al. 2002)
pDEST15	GST	pBR322	26.2	T7	non	Amp^R	(Walhout, Temple et al. 2000)
pDEST16	Thioredoxin	pUC	14.2	T7	Enterokinase	Amp^R	(Walhout, Temple et al. 2000)
pDEST17	6xHis	pBR322	2.4	T7	non	Amp^R	(Walhout, Temple et al. 2000)

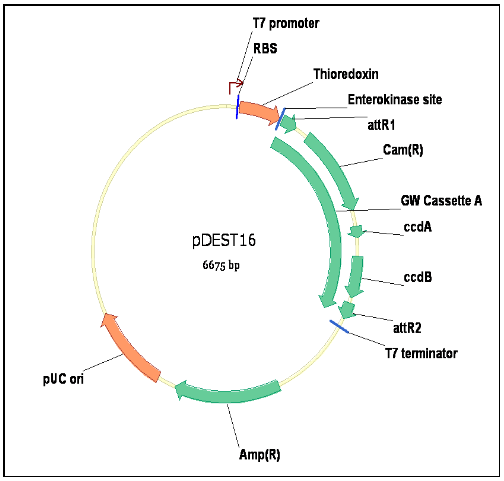

lacI
pUC ori
Lac operator
Tac promoter
pTH1
8525 bp
MBP
Factor Xa site
attR1
Amp(R)
Cam(R)
GW Cassette C
attR2
ccdA
ccdB


T7 terminator
attR2
ccdB
ccdA
Amp(R)
GW Cassette C
Cam(R)
attR1
TEV proteas site
Gb1
RBS
T7 promoter
Lac operator
pTH3
7556 bp
pUC ori
LacI


T7 promoter
RBS
GST
3C site
attR1
Cam(R)
pTH6
7025 bp
GW Cassette B
ccdA
ccdB
pUC ori
attR2
T7 terminator
Amp(R)


T7 terminator
attR2
ccdB
Carb(R)
ccdA
GW Cassette B
Cam(R)
pUC ori
attR1
3C site
pTH7
8637 bp
NusA
RBS
T7 promoter
Lac operator
LacI


T7 terminator
attR2
ccdB
ccdA
GW Cassette B
Amp(R)
Cam(R)
attR1
3C site
Z
T7 tag
RBS
T7 promoter
Lac operator
pTH10
7351 bp
pUC ori
LacI


T7 promoter
RBS
Thioredoxin
Enterokinase site
attR1
Cam(R)
GW Cassette A
ccdA
pDEST16
6675 bp
ccdB
attR2
T7 terminator
pUC ori
Amp(R)

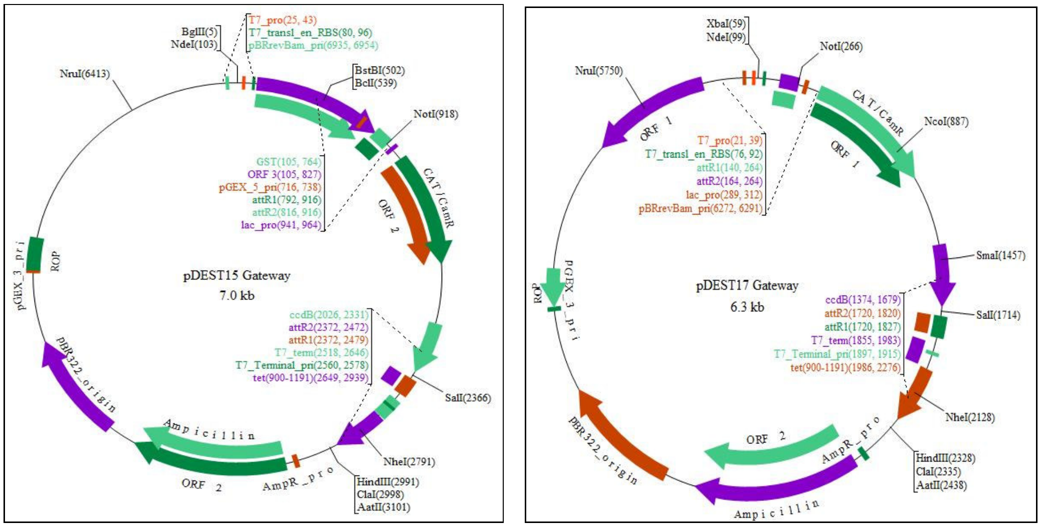


Figure 3.5.0: Expression vectors used to clone LexA for increased soluble yield.

3.5.1. Expression clone with 8 different vectors:

Cloning into above describe vectors were done as describe in section 3.28, 3.29 (LR cloning = BP clone x vector). Only one cell type was used for this system namely the BL21 (DE3) Codon Plus lysogenic cells, with an extra plasmid for rare codon supplementation. True clones were detected by colony PCR with gene specific primer.

3.5.2. Expression conditions:

For expression studies standard 2xYT medium was used with recommended concentrations (100µg/L ampicillin, 50 µg/L carbencillin) for the maintenance of vector containing strains. An overnight culture grown at 30 ºC was used to inculcate fresh 2xYT medium in the ratio of 1:10. Test expressions were done with 10ml 2xYT medium in 100ml erlenmeyer flasks at 220 rpm, for 3hrs at 37ºC.

3.5.3. Solubility Screening:

At the harvest point OD_{600} was determined and well mixed, 5ml of culture was taken from that to spin down. All the cultures were calculated for the resuspension volume according to (La Vallie 1993). Formula for determining volume for resuspension of cell pellet is given as under.

$$V_{\text{resuspension}} = (OD_{600}\text{ of culture}/ 5) \times \text{volume of culture}$$

3.6.0 Enzyme Assay for TVS4041:

Initial finding for the characterization of TVS4041 was brought about by checking the activity against 4 different chromogenic substrates. These substrate are known for detection of different type of serine protease activity. They are listed below:

- Succinyl-Ala-Ala-Pro-Ala-*p*-nitroanilide (AAPA) [for Elastase Activity]
- Succinyl-Ala-Ala-Pro-Leu-*p*-nitroanilide (AAPL) [for Elastase & Chymotrypsin]
- Succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide (AAPF) [for Chymotrypsin Activity]
- N-benzoyl-arginine-*p*-nitroaniline (BAPNA) [for Trypsin Activity]

3.6.1 Materials:

Following material needed for activity:

- 100mM stock solution of all 4 substrates in DMSO.
- 20mM HEPES buffer (pH 5.0) + 20mM $CaCl_2$
- 20mM HEPES buffer (pH 6.0) + 20mM $CaCl_2$
- 20mM HEPES buffer (pH 7.0) + 20mM $CaCl_2$
- 20mM HEPES buffer (pH 8.0) + 20mM $CaCl_2$
- solution Trypsin Std. solution (20µg/ml in 0.001M HCl)

3.6.2 Principal of Assays:

The protease can hydrolyze substrates at the bond between peptide and the *p*-nitroaniline moieties hence release the chromophore *p*-nitroaniline that can be detected by yellow color development or by colorimetric analysis at 410nm.

3.6.3. Methods:

Substrate working concentration was prepared by adding 1µl of substrate into 100 µl of buffer. Hence final concentration of reaction buffer was 1mM substrate, 1% DMSO, 10mM HEPES buffer, 10mM $CaCl_2$. 96 well plates were used to carryout the enzyme assay in crude extract of soluble protein.

3.7.0 Purification attempts for TVS4041:

3.7.1 Benzamidine Sepharose™ 6B (GE Health Care) Column:

Purification attempt was done initially through Benzamidine Sepharose column. *p*-Amino Benzamidine is synthetic inhibitor of trypsin and trypsin like Serine proteases. In this column *p*-amino benzamidine is covalently linked with 6% agarose. Maxium capacity of column was 13 mg/ drained medium, pressure limits were 0.3 Pa and pH range was 3-11 (GE 2006).

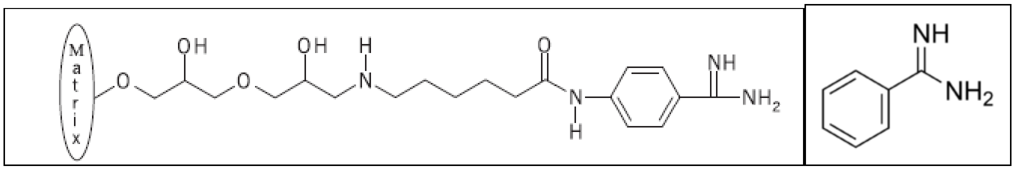


Figure 3.7.1: (a) partial structure of Benzamidine attached with agarose (b) benzamidine

3.7.2. Benzamidine column purification:

Periplasmic fraction was prepare through the procedure describe by pBAD/gIII (Invitrogen®) manual, applied onto the Benzamidine column 1 ml 6B (GE Gealthcare™). This fraction contains 20mM Tris-HCl pH 8.0 and 2.5mM EDTA it was further added with activation solution containing 25mM HEPES pH 7.5, 5M NaCl, 0.1M $CaCl_2$, in the ratio of 1:10ml of periplasmic sample.

Protein purification was performed through Äkta FPLC™ (Amershram Biopharmechia®). Running Buffer A contained 25mM HEPES, 10mM $CaCl_2$**,** Buffer B or Elution Buffer contained 25mM HEPES, 10mM $CaCl_2$, and 1M NaCl. Total 40 ml of culture was applied onto the column with the pressure limits 0.3 Mpa throughout the procedure and the flow rate at 0.5ml/min.

3.7.3. HisTrap HP™ column (GE Health Care):

HisTrap™ column design is based on the immobilized metal ion chromatography (IMAC). Column had less binding capability with biological molecules and minimal nickel ion leaking in allowed conditions. It selectively binds to any suitable complex forming amino acid (preferably poly Histidine) on the protein surface and can be eluted competitively by the elution with imidazole, which has higher affinity for nicle ion then poly histidine.

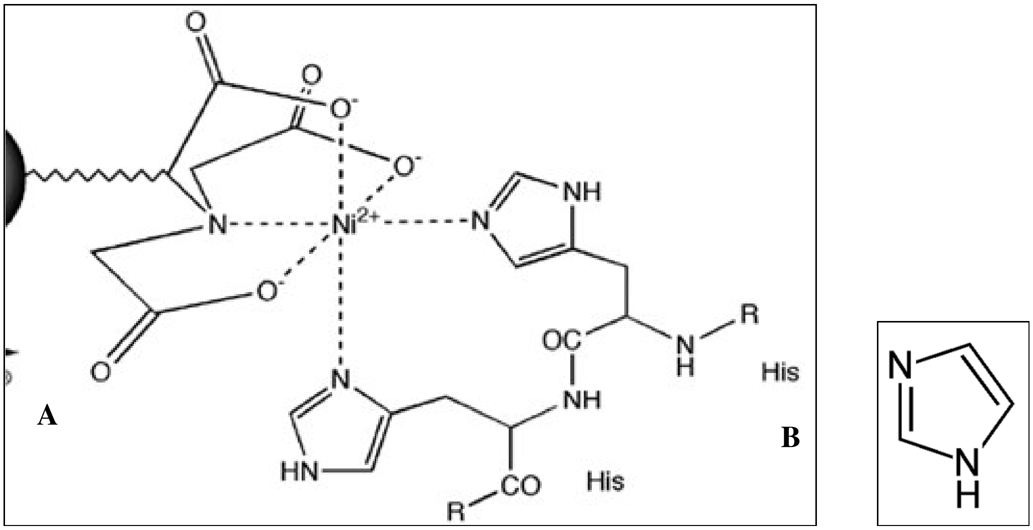


Figure 3.7.3.A: Poly His tail binding of His tagged protein to Ni-sephrose column. Figure A: Imidazole molecule

3.7.4. Recharging column

Stripping and recharging has done for higher efficiency in affinity purification, after every 3-4th time reuse of HisTrap column. Stripping buffer contained 50mM HEPES, pH 7, 0.3M NaCl, 0.05M EDTA. Stripping was done by 3-4 column washed with stripping buffer, followed by 3 column washed with Milli-Q water. Charging was done with 0.1M $NiCl_2$ in 2-3 column volumes followed by 2 columns washed with Milli-Q water.

3.7.5. His trap column Purification:

Periplasmic fraction obtain through the procedure described by pBAD/gIII (Invitrogen®) manual were further concentrated in spin filter concentrator (cut off 10KD MW) in a centrifuge at 45000rpm for 45 mins. The idea was to reduce the concentration of EDTA that was included in the periplasmic fraction in the concentration of 2.5mM and can cause the leaching of the Ni^{++} ion. So, a 30 ml of periplasmic sample was concentrated to the 1.2 ml volume in the above described procedure and then made-up the volume up to 5 ml with Buffer A. Protein purification was performed on ready to use HisTrap HP 1ml column (GE Healthcare™), through Äkta FPLC™ (Amershram Biopharmechia®) on 4°C refrigerated cabinet. **Running buffer A** contained 25mM HEPES, 10mM $CaCl_2$, 1% glycerol, **buffer B** or **elution buffer** contained 25mM HEPES, 10mM $CaCl_2$, 1% glycerol and 500mM imidazole.

.

Results and Discussion

4.1.0. Degradomic analysis of *Aliivibrio salmonicida* genome:

Aliivibrio salmonicida genome consists of two circular chromosomes (chrI 3.3 Mb, chrII 1.2Mb) and additionally four circular plasmids designated pVSAL840 (83.5kb), pVSAL320 (30.8 kb), pVSAL54 (5.4 kb) and pVSAL43 (4.3 kb) which represent 2.7% of the total genomic DNA. As a whole total genome consist of 4.6 Mb, encoding for 4,286 proteins (Hjerde, Lorentzen et al. 2008). The genome was found to consist of roughly 113 protease activity containing proteins and many protease inhibitors where a criterion of significance was set out with the e-score <1 (Altschul and Lipman 1990; Altschul, Wootton et al. 2005). Details of these proteases are given in section 7 (appendices), table 7.1. Brief description of proteases from *A. salmonicida* and other vibrio species are given in table 4.1 for comparision.

Table 4.1: comparision of protease contants in evolutionary closer species of *Aliivibrio salmonicida,* where A, C, G, M, T, S, U stands for Asp, Cys, Glu, Metallo, Thr, Ser, Unassigned type of proteases alternativly. Source; MEROPS database

Organism	Total proteases	A	C	G	M	T	S	U
'Vibrionales bacterium SWAT-3'	44	-	-	-	-	-	-	-
Vibrio shilonii	46	-	-	-	-	-	-	-
Vibrio sp. Ex25	46	-	-	-	-	-	-	-
Vibrio sp. MED222	58	-	-	-	-	-	-	-
Vibrio angustum	69	-	-	-	-	-	-	-
Vibrio alginolyticus	72	-	-	-	-	-	-	-
Vibrio harveyi	82	-	-	-	-	-	-	-
Vibrio fischeri	97	4	6	0	35	2	43	7
Vibrio vulnificus	98	-	-	-	-	-	-	-
Vibrio parahaemolyticus	99	-	-	-	-	-	-	-
Pseudoalteromonas haloplanktis	104	-	-	-	-	-	-	-
Aliivibrio salmonicida	113	5	5	1	48	3	41	10
Vibrio cholerae O1, biovar eltor	115	6	14	0	43	4	42	6
Vibrio splendidus	128	-	-	-	-	-	-	-

From the detected proteases 1 was Glutamic type; 3 were from Thereonine type; 5 were from Cystine type; 5 were Aspartic type; 41 from Serine type, 48 were from Metallo type, and 10 were from Unknown type of peptidase families. This represents 2.6% of total ORFeome contents of *Aliivibrio salmonicida.* Among all proteases metallo type protease contribute almost the major content (42.48% of identified protease). This could be related to the virulence character of the organism, since metallo type proteases are capable of major tissue degradation, as a first step in host body invasion (Hjerde 2007).

4.2.0. Target selection:

The following targets were selected from the whole proteome pool on the bases of interesting features. Following are their biological functions and protease type reported in MEROPS database (www.**merops**.sanger.ac.uk).

- **LexA homolog:** LexA is a repressor protein for more than 20 SOS genes, involved in survival of DNA-damaged cells. The single stranded DNA appears in the vicinity of damaged DNA initiate RecA dependent autoclavage of LexA, resulting in release of DNA repairing genes to be transcribed.
- **RadA homolog:** RadA is a DNA repairing protein. Study of this protein, was interesting to evaluate the function of this protein in *A. salmonicida.*
- **ThiJ/PfpI homolog:** Cystine type of protease. Function predicted as isoprenoid biosynthesis protein with amidotransferase-like domain. Homologue found in hexameric ring structure. Unique for the study of oligomerization and function.
- **HslV homolog:** Threonine type of peptidase. Prokaryotic type of proteosom subunit. Some time refers as heat shock protein. Unique in a sense of catalytic type and the absence of high resolution structural information of homolog from *E. coli.*
- **Trypsin homolog:** Serine type of protease S1/S6 family. Predicted to have the chymotrypsin-like proteolytic activity. Important due to commercial potentials.
- **Collagnase homolog:** Metallo type of protease enzyme, from M9 family. Important due to the commercial potentials.
- **ProtinaseK homolog:** serine type of degradative protease, from S8 family. Important due to the commercial potentials.
- **ToxR homolog:** Metallo type protease activity containing protein, reside transmembrane structure. Unique for study since no homolog was found in PDB database.

Table 4.2.0: Bioinformatic analysis of selected targets for expression.

ORF A.A	MEROPS	(Blast with NCBI nr Data Base)						CD Blast		Blast with PDB	
		Highest similarity found with									
	Peptide Family	Protein	Organism	% Id	%+Ve	Score	E-value	code	% aligned	PDB code	%Positive
207	A8	LexA repressor	*Vibrio fischeri* ES114	190/208 (91%)	199/208 (95%),	322	5e-87	CDD|11682	99.0	1JHH	173/207 (83%)
459	S16	DNA repair protein RadA	*Vibrio fischeri* ES114	441/460 (95%)	453/460 (98%)	868	0.0	CDD|10790	100.0	1RR9	49/89 (55%)
218	C56	ThiJ/ PfpI	*Pseudomonas syringae*	109/212 (51%)	152/212 (71%),	234	2e-60	CDD|10562	98.2	1OY1	142/212 (66%)
180	T1B	HslV	*Photor-habdus luminescens*	124/172 (72%)	144/172 (83%)	236	4e-61	CDD|30160	100.0	1NED	143/171 (83%)
323	S1A	Trypsin	*Vibrio fischeri* ES114	191/327 (58%)	234/327 (71%)	352	2e-95	CDD|29152	98.7%	1PYTD	119/263 (45%)
824	M9	Collagnase	*Vibrio fischeri* YJ016	366/808 (45%)	528/808 (65%)	717	0.0	CDD|23266	100.0	1JLR	19/27 (70%)
479	S8	ProtinaseK	*Photobacteri-um profundum* SS9	390/480 (81%),	432/480 (90%),	776	0.0	Pfam|00082	88.2	1DBI	105/261 (40%)
287	M23B	ToxR activating protein	*Vibrio fischeri* ES114	252/287 (87%)	276/287 (96%)	497	3e-139	CDD|2090	100.0%	2BI3	46/86 (53%)

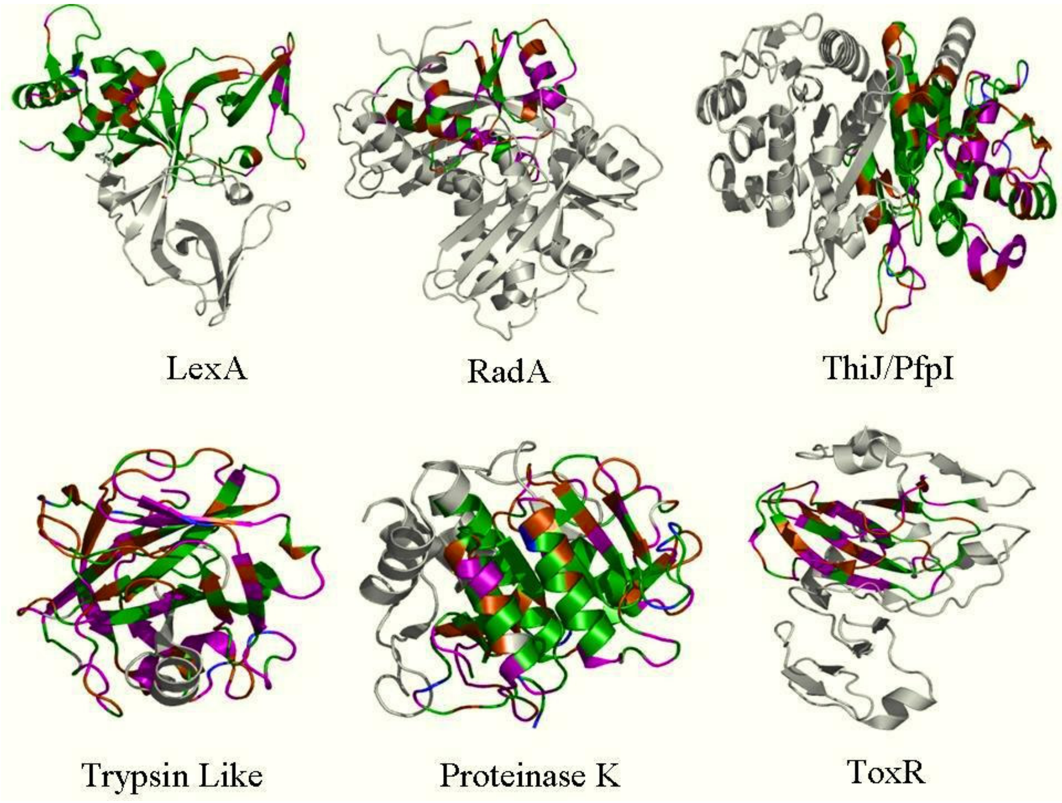


Figure 4.2.0: Model formation through FeatureMap3D http://www.cbs.dtu.dk/services/FeatureMap3D/.

4.3.0 Analysis of TVS4041:

Detail bioinformatic study of TVS4041 is given here, since this target protein was under major investigation. Several factors have been discussed that have major impact expression, purification and characterization of this target protein.

4.3.1. GC:AT ratio analysis:

The percentage of AT and GC calculated through NEB cutter analysis (http://tools.neb.com/NEBcutter2/index.php), was GC=42%, AT=58%. This represents a better situation in recombinant expression since, AT contants are not much higher than GC contants.

4.3.2. PI and molecular weight of TVS4041 after cloning:

The sequence in bold alphabets (Pink) represents protein's own sequence, while non-bold alphabets (green) represent linker sequence from vector.

N-terminal, with pro sequence:	ACC ATG **GCA ACC GAA GAG TTT TCA**
	Thr Met **Ala Thr Glu Glu Phe Ser**
N-terminal, without Pro sequence:	ACC ATG **GGC ATT GTG GGT GGT AAT**
	Thr Met **Gly Ile Val Gly Gly Asp**
C-terminal Native:	**AGA GCG TCA TGA** GTC
	Arg Ala Ser stop Val
C-terminal His:	**TCA** GTC GAC ATT 6(CAT) TGA GTT
	Ser Val Asp Ile 6His stop Val

>TVS4041_recombinant_N1 (MW 32.5, PI 4.65)

TMATEEFSVTPYIVGGNDANVAGYPFMASLMFEYASQPGVIYPFCGGSILDSTHILTAAHCVY
DTASSQVSNMKVAIEANNGQGMLAAQRVAVKNIYYPSDYNDSTLINDVEVLELSEALPNYTL
GHAATLGESYLEGQGYRAVGSIFTIIGYGRLSSTQANTNVDFMEARVKYVNPTDCNVWANFT
TSDKQVCSSGYSFDSSDLVTATCQGDSGGPLVWNGTQIGIVSFGPSVCGQTIVATGTLPAQSVF
TDVSHYKDWILKAQRGEVTSTITATTSSSSSGGSIPLFGLLGLSLFGYYRKSRAS

>TVS4041_recombinant_N2 (MW 31.37, PI 4.77)

TMIVGGNDANVAGYPFMASLMFEYASQPGVIYPFCGGSILDSTHILTAAHCVYDTASSQVSNM
KVAIEANNGQGMLAAQRVAVKNIYYPSDYNDSTLINDVEVLELSEALPNYTLGHAATLGESY
LEGQGYRAVGSIFTIIGYGRLSSTQANTNVDFMEARVKYVNPTDCNVWANFTTSDKQVCSSG
YSFDSSDLVTATCQGDSGGPLVWNGTQIGIVSFGPSVCGQTIVATGTLPAQSVFTDVSHYKDWI
LKAQRGEVTSTITATTSSSSSGGSIPLFGLLGLSLFGYYRKSRAS

>TVS4041_recombinant_H1 (MW 33.8, PI 5.11)

TMATEEFSVTPYIVGGNDANVAGYPFMASLMFEYASQPGVIYPFCGGSILDSTHILTAAHCVY
DTASSQVSNMKVAIEANNGQGMLAAQRVAVKNIYYPSDYNDSTLINDVEVLELSEALPNYTL
GHAATLGESYLEGQGYRAVGSIFTIIGYGRLSSTQANTNVDFMEARVKYVNPTDCNVWANFT
TSDKQVCSSGYSFDSSDLVTATCQGDSGGPLVWNGTQIGIVSFGPSVCGQTIVATGTLPAQSVF
TDVSHYKDWILKAQRGEVTSTITATTSSSSSGGSIPLFGLLGLSLFGYYRKSRAS**ASVDIHHHH
HH**

>TVS4041_recombinant_H2 (MW 32.68, PI 5.32)

TMIVGGNDANVAGYPFMASLMFEYASQPGVIYPFCGGSILDSTHILTAAHCVYDTASSQVSNM
KVAIEANNGQGMLAAQRVAVKNIYYPSDYNDSTLINDVEVLELSEALPNYTLGHAATLGESY
LEGQGYRAVGSIFTIIGYGRLSSTQANTNVDFMEARVKYVNPTDCNVWANFTTSDKQVCSSG
YSFDSSDLVTATCQGDSGGPLVWNGTQIGIVSFGPSVCGQTIVATGTLPAQSVFTDVSHYKDWI
LKAQRGEVTSTITATTSSSSSGGSIPLFGLLGLSLFGYYRKSRAS**ASVDIHHHHHH**

4.3.3. Rare codon:

Rare codon were calculated using http://nihserver.mbi.ucla.edu/RACC/

Red = rare Arg codons AGG, AGA, CGA
Green = rare Leu codon CTA

```
gag gtt tta ctc gtg act aca att gtt tct gtt cgc cgt gaa
ggc aaa gtc gtt att gct ggt gat ggc caa gca tct caa ggc
gat atg atc gct aaa ggc aac gta aaa aaa gtt cgt cgt tta
tat aac gat tct gta ctc gtt gga ttc gca ggc agc acc gca
gat gcc ttc att ttg ttc gac CTA tgt gaa CGA aaa tta gaa
atg cac caa ggt aat tta acg aaa gcc gcc gtt gaa CTA gca
aaa gat tgg cgt agt gat cgt aat tta cgt cgt ctt gaa gcc
atg CTA att gtt gcc gat gac act acc tca ctg atc atc agt
ggt act ggt gac tta atc aat gca gat aat gac ctc ctc act
att ggt tct ggt ggt tat ttt gct cgt tct gcg gca acc gca
tta tta gaa aat aca gat tta gat gca tac gat att gca gtg
aaa gca ctg act atc gct ggt gat act gat gta tac acc aat
cat aac cat acc gtt gaa gtt ctt gat acc aac aaa tag
```

Frequency of occurance of rare codones in TVS4041 is not very high based on intrepration. But use of *E. coli* strains that bear the plasmid for coding tRNA for CGA (Arg) and CTA (Leu) is prefential, like Rosetta branded cells that bear plasmid encoding both of these rare codons.

4.3.4. Estimated half-life:

The Protparam tool (ExPASY) was used to estimate the half life of the enzyme. For native TVS4041 (with IVGG in *N*-terminal), it was estimated to be for 20 Hrs in mammalian reticulocytes (*in vitro*), 30 mins in yeast (*in vivo*), and 10 hrs in *E. coli* (*in vivo*). While with recombinant sequences (where TM is used in *N*-terminal) it is calculated as 7.2 hrs in mammalian reticulocytes (*in vitro*), >20 hrs in yeast (*in vivo*), and >10 hrs in *E. coli* (*in vivo*). Hence it can be seen that due to the linker region half life of this target protein has now expected to be increased 19 times in yeast cells. While in *E. coli*, it is expected to be stable for at least more than 10hrs. Therefore, pilot expression studies, of this particular protein was decided to perform in *E. coli*.

4.3.5. Instability index:

This theoretical value is based on the frequency of 400 studied instability weight value dipeptides occurrence in 12 unstable and 32 stable proteins (Guruprasad, Reddy et al. 1990). According to the protparam tool; instability index (II) is computed to be 28.33 for native sequence, 29.43 for N1, 28.21 for N2, 28.67 for H1, 27.47 for H2, which classifies the protein as stable, since value above 40 will be considered as unstable according to experimental evaluation.

The formula of instability is given as following:

$$\mathrm{II} = \frac{10}{\mathrm{L}} \times \sum_{\mathrm{i=1}}^{\mathrm{i=L\text{-}1}} \mathrm{DIWV}(\mathrm{x}[\mathrm{i}]\mathrm{x}[\mathrm{i}+1])$$

Where L is the length of sequence and DIWV (x[i]x[i+1]) is the instability weight value for the dipeptide starting in position i.

4.3.6. Aliphatic index:

The thermostability of globular proteins depends on the relative volume occupied by side chains of aliphatic residues (alanine, valine, isoleucine, and leucine) in globular proteins (Ikai 1980). For TVS 4041 native was estimated as 81.26 for N1

79.35, for N2 80.71, for H1 79.05, for H2 80.36 percent mole fraction. This stability index, predict the ability of TVS4041, to bear moderately higher temperature.

The formula for estimation is as following:

$$\text{Aliphatic index} = X\ (\text{Ala}) + a * X\ (\text{Val}) + b * [X\ (\text{Ile}) + X\ (\text{Leu})]$$

Where X(Ala), X(Val), X(Ile), and X(Leu) are mole percent (100 X mole fraction) of alanine, valine, isoleucine, and leucine. The coefficients *a* and *b* are the relative volume of valine side chain (a = 2.9) and of Leu/Ile side chains (b = 3.9) to the side chain of alanine.

4.3.7. Grand average of hydropathicity (GRAVY):

The hydrophobicity calculated on the basis of Kyte & Doolitlee scale was estimated to be -0.019 for TVS 4041, which is the sum of all the hydropathy values for individual residues divided by number of residues (Kyte and Doolittle 1982).

$$\text{GRAVY} = \frac{\sum(A + C + I + L + M + F + V) - (R + N + D + Q + E + G + H + K + P + S + T + W + Y)}{N}$$

The calculated value of '-0.019' indicates slight hydrophilicity of the whole polypeptide. It means on average this protein will be tended to be soluble in hydrophilic solution.

4.3.8. Insolubility analysis "recombinant solubility prediction":

Table 4.3.8 : Insolubility analysis for TVS4041 by "recombinant solubility prediction" (Wilkinson and Harrison 1991). Underlined entries (brown color) denote the PDB entries of full functional domains that aligned with the chopped *C*-terminal domain of TVS4041.

Sequence code/ detail	Sequence domains	Predicted state
E0 (Signal sequence included)	MNVVVGALVSLSLLSPVVL ATEEFSVTPY IVGGNDANVAGYPFMASLMFEYASQPGVIYPFCGGSILDSTHILTAA HCVYDTASSQVSNMKVAIEANNGQGMLAAQRVAVKNIYYPSDYND STLINDVEVLELSEALPNYTLGHAATLGESYLEGQGYRAVGSIFTIIG YGRLSSTQANTNVDFMEARVKYVNPTDCNVWANFTTSDKQVCSSG YSFDSSDLVTATCQGDSGGPLVWNGTQIGIVSFGPSVC GQTIVATGTLPAQSVFTDVSHYKDWIL KAQRGEVTST ITATTSSSSS GGSIPLFGLLGLSLFGYYRKSRAS	74.4% InS
E1 (signal sequence chopped)	ATEEFSVTPY IVGGNDANVAGYPFMASLMFEYASQPGVIYPFCGGSILDSTHILTAA HCVYDTASSQVSNMKVAIEANNGQGMLAAQRVAVKNIYYPSDYND STLINDVEVLELSEALPNYTLGHAATLGESYLEGQGYRAVGSIFTIIG YGRLSSTQANTNVDFMEARVKYVNPTDCNVWANFTTSDKQVCSSG YSFDSSDLVTATCQGDSGGPLVWNGTQIGIVSFGPSVC GQTIVATGTLPAQSVFTDVSHYKDWIL KAQRGEVTST ITATTSSSSS GGSIPLFGLLGLSLFGYYRKSRAS	73.1% InS
E2 (pre and pro sequence chopped)	IVGGNDANVAGYPFMASLMFEYASQPGVIYPFCGGSILDSTHILTAA HCVYDTASSQVSNMKVAIEANNGQGMLAAQRVAVKNIYYPSDYND STLINDVEVLELSEALPNYTLGHAATLGESYLEGQGYRAVGSIFTIIG YGRLSSTQANTNVDFMEARVKYVNPTDCNVWANFTTSDKQVCSSG YSFDSSDLVTATCQGDSGGPLVWNGTQIGIVSFGPSVC GQTIVATGTLPAQSVFTDVSHYKDWIL KAQRGEVTST ITATTSSSSS GGSIPLFGLLGLSLFGYYRKSRAS	77.2% InS
C1 (C-terminal chopped) Delta chymotrypsin Pdb:1dlk/B	IVGGNDANVAGYPFMASLMFEYASQPGVIYPFCGGSILDSTHILTAA HCVYDTASSQVSNMKVAIEANNGQGMLAAQRVAVKNIYYPSDYND STLINDVEVLELSEALPNYTLGHAATLGESYLEGQGYRAVGSIFTIIG YGRLSSTQANTNVDFMEARVKYVNPTDCNVWANFTTSDKQVCSSG YSFDSSDLVTATCQGDSGGPLVWNGTQIGIVSFGPSVC	62.6% InS
C2 (C-terminal chopped) Human Chymase Pdb:1pjp	ATEEFSVTPY IVGGNDANVAGYPFMASLMFEYASQPGVIYPFCGGSILDSTHILTAA HCVYDTASSQVSNMKVAIEANNGQGMLAAQRVAVKNIYYPSDYND STLINDVEVLELSEALPNYTLGHAATLGESYLEGQGYRAVGSIFTIIG YGRLSSTQANTNVDFMEARVKYVNPTDCNVWANFTTSDKQVCSSG YSFDSSDLVTATCQGDSGGPLVWNGTQIGIVSFGPSVC GQTIVATGTLPAQSVFTDVSHYKDWIL	53.4% InS
C3 (C-terminal chopped) Snake venome Pdb:1OP2/A	IVGGNDANVAGYPFMASLMFEYASQPGVIYPFCGGSILDSTHILTAA HCVYDTASSQVSNMKVAIEANNGQGMLAAQRVAVKNIYYPSDYND STLINDVEVLELSEALPNYTLGHAATLGESYLEGQGYRAVGSIFTIIG YGRLSSTQANTNVDFMEARVKYVNPTDCNVWANFTTSDKQVCSSG YSFDSSDLVTATCQGDSGGPLVWNGTQIGIVSFGPSVC GQTIVATGTLPAQSVFTDVSHYKDWIL KAQRGEVTST	61.9% InS
C4 (C-terminal chopped) Beta Acrosin from Ram spermatozoa Pdb:1fiw/A	IVGGNDANVAGYPFMASLMFEYASQPGVIYPFCGGSILDSTHILTAA HCVYDTASSQVSNMKVAIEANNGQGMLAAQRVAVKNIYYPSDYND STLINDVEVLELSEALPNYTLGHAATLGESYLEGQGYRAVGSIFTIIG YGRLSSTQANTNVDFMEARVKYVNPTDCNVWANFTTSDKQVCSSG YSFDSSDLVTATCQGDSGGPLVWNGTQIGIVSFGPSVC GQTIVATGTLPAQSVFTDVSHYKDWIL KAQRGEVTST ITATTSSSSS	65.9% InS
Pro sequence	ATEEFSVTPY	97% Sol
Tail sequence	KAQRGEVTST ITATTSSSSS GGSIPLFGLLGLSLFGYYRKSRAS	93.4% InS
	ITATTSSSSS	97% InS

This analysis revels the presence of highly insoluble structure in the *C*-terminal of TVS4041 while the pro sequence is oppositely highly soluble segment in this protein. Among the analyzed chopped structures that can be functional after the removal of segments **C2** has been found to be the most soluble structure that is predicted to improve the solubility and to yield a full functional form. Interestingly, the most influencing part on this protein is in *N*- and *C*-terminals, and according to the known structure it seems they will be exposed on the surface of this protein and can influence the solubility and yield.

In connection to the accusation of faulty pridiction from this kind of softwares for certain proteins are due to their mean evolution from exposed and unexposed surfaces residues of protein. While the surface residues mostly account for insolubility due to their hydrophobicity on the surface and inability to stand in hydrophilic environment.

4.3.9. The Amino Acid composition of mature full length TVS4041 (native):

From the table 4.3.9, it can be seen that the highest number of amino acid residues are of Ser, Gly, Ala, Thr and Val that account for 47 % of total residues. The common features of them are their smaller size. Among them highest number is of Ser, which is known to confer the thermostability by forming additional intramolecular hydrogen bonds, such as observed in critical places of triosephosphate isomerase (Alvarez, Zeelen et al. 1998).

The second highest ratio of Gly represents the structural flexibility contents. Since, Gly do not have any side chain and it can adapt many conformational directions. In contrast, Pro confers rigidity in the structure due to the rotational prohibition around Φ angle. This amino acid is also considered important in controling the protein folding *in vivo* (Branden and Tooze 1999). TVS4041 do not have very higher contants of Pro to Gly ratio (3.4: 10.2), indicating a fexible structure that could folde in higher rate in absence of less Pro contents.

Numbers of negatively charged residues (Asp + Glu) are 7.9%, while numbers of positively charged residues (Arg + Lys) are 4.8%. Hence, total charged residues (RLHED) are 14.1% of the whole polypeptide. Calculated net charge (RK-DE) is estimation to -3.1%, that is close to the average trypsins charges (Leiros, Willassen et al. 1999).

Aromatic residues (FWY) are 10.1% in comparison to aliphatic residues (GILV) 30.9%. Hydrophobic residues are 39% in comparision to 61% of hydrophilic residues (QNCSTYRKHDEG).

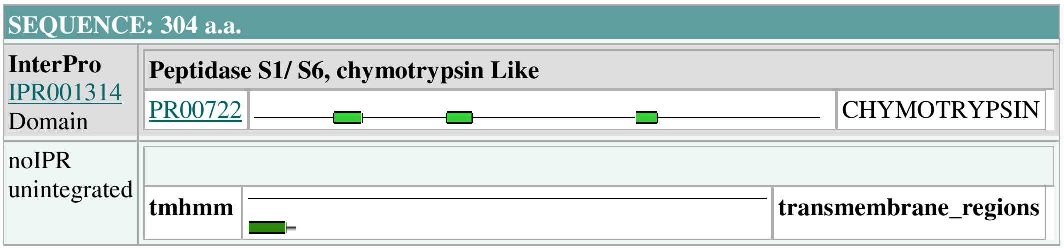
Table 4.3.9: The amino acid composition of full length, TVS4041.

Amino acid	quantity	percnetage
Ser (S)	35	11.9%
Gly (G)	30	10.2%
Ala (A)	25	8.5%
Thr (T)	24	8.2%
Val (V)	24	8.2%
Leu (L)	21	7.1%
Ile (I)	16	5.4%
Tyr (Y)	16	5.4%
Asn (N)	15	5.1%
Asp (D)	14	4.8%
Gln (Q)	12	4.1%
Phe (F)	11	3.7%
Pro (P)	10	3.4%
Glu (E)	9	3.1%
Lys (K)	7	2.4%
Arg (R)	7	2.4%
Cys (C)	6	2.0%
His (H)	4	1.4%
Met (M)	5	1.7%
Trp (W)	3	1.0%

4.4.1. Catalytic type of TVS4041:

When blasting into InterProScan (www.ebi.ac.uk /Tools) the catalytic type was predicted to be a serine type of protease. Serine proteases have three broader categories Chymotrypsin like, Subtilisin and Carboxypeptidase C, all three have serine, aspartate and histidine in common. Geometric orientation of these residues, are common but arrangement order of these amino acids on a continuous polypeptide is amazingly different. The catalytic triad in the Trypsin like domain is ordered HDS, and DHS in the subtilisin and SDH in the carboxypeptidase. In TVS4041 'domain analysis' reveals the presence of active sites in such order that proves it to be of trypsin/chymotrypsin-like family. This prediction also indicates presence of transmembrane segment in the extra long *C*-terminal of TVS4041 (Table 4.4.1).

Table 4.4.1: The domain evaluation of TVS4041, with **InterProScan**.



SEQUENCE: 304 a.a.			
InterPro IPR001314 Domain	**Peptidase S1/ S6, chymotrypsin Like**		
	PR00722		CHYMOTRYPSIN
noIPR unintegrated			
	tmhmm		**transmembrane_regions**

Chymotrypsin-like family is further divided in several branches on the bases of function like digestive enzymes, coagulation factors, kallikrein, cathapsinG, enterokinases, growth factor activators, hepsins, tryptases, snake venoms, collagnases, allergens etc.

Most serine proteins from lower organism either play function in digestion or virulency that results in pathogenic conditions or allegic reactions in the hosts. In lower organisms these trypsin-like domain are divided in four broader categories on the bases of their substrate specificity. These are chymotrypsin, trypsins, elastase and collagenase. In discussion with the category that can be assigned to TVS4041, further features needed to be evaluating by alignment, phylogenetic tree formation and model building. During the alignment several other serine proteases were observed that were closed homolog to TVS4041, but they were specialized for function only in higher organisms in certain aspects like, kallikrein, cathapsinG, coagulation factors, Factor Xa, Thrombin etc.

4.4.2. Sequence alignment and comparisions:

The amino acid sequence of TVS4041 was used for Blast search (BLASTp) (Altschul, Gish et al. 1990) using both 'non-redundant' amino acid sequence, and structurally resolved PDB database (blast date:12th, May 2009). The retrived sequences were then aligned using ClustalW (Thompson, Higgins et al. 1994; Thompson, Gibson et al. 1997; Thompson, Gibson et al. 2002).

The initial alignment was adjusted using BioEdit and GenDoc (Hall 1997; Nicholas and Nicholas 1997). Source: [http://www.mbio.ncsu.edu/BioEdit/bioedit.html, www.nrbsc.org/gfx/**genedoc**/index.html].

The alignment describe in Figure 4.4.2.1 is to identify the essential residues in connections with well known and well studied chymotrypsin family protease. Numbering was adapted on the bases of chymotrypsinogen A (Hrtley and Kauffman 1996). Secondry structure designation based on the published work from our group (Leiros, Willassen et al. 1999). From this alignment it is obvious that TVS4041 is much different to the conserved similar sequences from vertebral sources. This alignment will be discussed more in detail, when comparing the different type of trypsin structures and TVS4041.

Except alignment in figure 4.2.2.1, three more alignments were performed to evaluate the pattern of sequence variation in closely linked *vibrio spp.* presented in figure 4.4.2.2. This alignment was based on the blast results retrived from non redundant database. The second alignment was made with closer homologs from PDB datbase only belonging to vertebral sources figure 4.4.2.3. The third alignment includes homologs from invertebral sources includes insectal and microbial serine protease, represented in figure 4.4.2.4.

Figure 4.4.2.1: Alignment table for identification of key residues, active site residues (red), disulphide bonding cystine (yellow), substrate specificity determing residues (green), aligned residues (light blue), residues varying and inserted in specificity pocket (pink), the reported sites for autocatalysis are shown in red fonts. where, the sequence labeled with pdb codes, such as 1BTP is Bovine Trypsin; 1DPO is Rat Trypsin; 1A0J is ColdFish Trypsin; 1HJ8 is Salmon Trypsin.

	N- terminal							N- terminal loop											Nβ1								Nβ1-Nβ2								Nβ2								Nβ2-Nβ3			Nβ3				Nβ3- Nβ4 loop											
	16	17	18	19	20	21	22	23				24	25	26		27	28	29	30	31	32	33				34		37			38			39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58		59	60			61	62	63
1BTP		V	G	G	Y	T	C	G	.	.	.	A	N	T	.	V	P	Y	Q	V	S	L	.	.	.	N	.	S	.	.	G	.	.	Y	H	F	C	G	G	S	L	I	N	S	Q	W	V	V	S	A	A	H	C	.	Y	K	.	.	S	G	I
1DPO		V	G	G	Y	T	C	.	Q	E	.	.	N	S	.	V	P	Y	Q	V	S	L	.	.	.	N	.	S	.	.	G	.	.	Y	H	F	C	G	G	S	L	I	N	D	Q	W	V	V	S	A	A	H	C	.	Y	K	.	.	S	R	I
1A0J		V	G	G	Y	E	C	.	R	K	.	.	N	S	.	A	S	Y	Q	A	S	L	.	.	.	Q	.	S	.	.	G	.	.	Y	H	F	C	G	G	S	L	I	S	S	T	W	V	V	S	A	A	H	C	.	Y	K	.	.	S	R	I
1HJ8		V	G	G	Y	E	C	.	K	.	.	A	Y	S	Q	.	P	H	Q	V	S	L	.	.	.	N	.	S	.	.	G	.	.	Y	H	F	C	G	G	S	L	V	N	E	N	W	V	V	S	A	A	H	C	.	Y	K	.	.	S	R	V
TVS4041		V	G	G	.	N	D	A	N	.	V	A	G	.	.	Y	P	F	M	A	S	L	M	F	E	Y	A	S	Q	P	G	V	I	Y	P	F	C	G	G	S	I	L	D	S	T	H	I	L	T	A	A	H	C	V	Y	D	T	A	S	S	.

	Nβ4					Calcium binding loop															Nβ5														Nβ5_Nβ6 loop															Nβ6					
	64	65	66	67		69	70	71	72	73			74	75	76	77	78		79	80		81	82	83		84	85	86	87	88	89	90			91	92	93		94	95	96	97		98	99	100	101	102	103	104		105	106	107	108
1BTP	Q	V	R	L	.	G	E	D	N	I	.	.	N	V	V	E	G	.	N	E	.	Q	F	I	.	S	A	S	K	S	I	V	.	.	H	P	S	.	Y	N	S	N	.	T	L	N	N	D	I	M	.	L	I	K	L
1DPO	Q	V	R	L	.	G	E	H	N	I	.	.	N	V	L	E	G	.	N	E	.	Q	F	V	.	N	A	A	K	.	I	I	.	K	H	P	N	.	F	.	D	R	K	T	L	N	N	D	I	M	.	L	I	K	L
1A0J	Q	V	R	L	.	G	E	H	N	I	.	A	V	N	.	E	G	.	T	E	.	Q	F	I	D	S	.	V	K	.	V	I	M	.	H	P	S	.	Y	N	S	R	.	N	L	D	N	D	I	M	.	L	I	K	L
1HJ8	E	V	R	L	.	G	E	H	N	I	K	.	.	V	T	E	G	.	S	E	.	Q	F	I	.	S	S	S	R	.	V	I	.	R	H	P	N	.	Y	S	S	Y	.	N	I	D	N	D	I	M	.	L	I	K	L
TVS4041	Q	V	S	.	N	M	**K**	V	A	I	E	A	N	N	G	Q	G	M	L	A	A	Q	.	.	R	V	A	V	K	N	I	Y	Y	.	.	P	S	D	Y	N	D	S	.	T	L	I	N	D	V	.	E	V	L	E	L

	Interdomain loop																																				Cβ1				Autolysis loop																		
	109	110					111	112	113	114	115	116	117	118	119	120	121		122		123		124	125	127	128	129		130	132		133				134	135	136	137	138	139	140	141	142	143	144	145		146	147	148	149	150	151			152	153	154
1BTP	K	S	.	.	.	.	A	A	S	L	N	S	R	V	A	S	I	.	S	.	L	.	P	T	S	C	A	.	S	A	.	G	.	.	.	T	Q	C	L	I	S	G	W	G	N	T	**K**	.	S	S	G	T	S	Y	.	.	P	D	V
1DPO	S	S	.	.	.	P	V	.	K	L	N	A	R	V	A	T	V	A	.	.	L	.	P	S	S	C	A	.	P	A	.	G	.	.	.	T	Q	C	L	I	S	G	W	G	N	T	.	L	S	S	G	V	N	E		.	P	D	L
1A0J	.	S	K	.	.	P	A	.	S	L	N	S	Y	V	S	T	V	A	.	.	L	.	P	S	S	C	A	S	S	.	.	G	.	.	.	T	R	C	L	V	S	G	W	G	N	L	S	G	S	S	S	N	.	Y	.	.	P	D	T
1HJ8	.	S	K	.	.	P	A	.	T	L	N	T	Y	V	Q	P	V	A	.	.	L	.	P	T	S	C	A	.	P	A	.	G	.	.	.	T	M	C	T	V	S	G	W	G	N	T	.	M	S	S	T	A	D	S			N	K	.
TVS4041	.	S	E	A	L	P	N	Y	T	L	G	H	A	A	T	L	G	E	S	Y	L	E	.	G	Q	.	G	Y	**R**	A	V	G	S	I	F	T	.	.	I	I	.	G	Y	G	.	.	**R**	L	S	S	T	.	Q	A	.	.	N	T	N

	Cβ2									Cβ2-Cα1				Cα1 helix							Cα1-Cβ3 loop							Cβ3							Loop1								Specificity pocket													Cβ4				Cβ4-Cβ5	
	155	156	157	158	159	160			161	162	163	164		165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180		181	182	183		184A	184B		185	186	187	188A	188B	189					190	191	192	193	194	195	196	197	198	199	200	201	202	203
1BTP	L	K	C	L	K	A	.	.	P	I	L	S	.	D	S	S	C	K	S	A	Y	P	G	Q	I	T	S	N	M	.	F	C	A	.	G	Y	.	L	E	G	G	K	D	.	.	.	.	S	C	Q	G	D	S	G	G	P	V	V	C	S	G
1DPO	L	Q	C	L	D	A	.	.	P	L	L	.	P	Q	A	D	C	E	A	S	Y	P	G	K	I	T	D	N	M	.	V	C	V	.	G	F	.	L	E	G	G	K	D	.	.	.	.	S	C	Q	G	D	C	G	G	P	V	V	C	N	G
1A0J	L	R	C	L	D	.	.	L	P	I	L	S	.	S	S	S	C	N	S	A	Y	P	G	Q	I	T	S	N	M	.	F	C	A	.	G	F	.	M	E	G	G	K	D	.	.	.	.	S	C	Q	G	D	S	G	G	P	V	V	C	N	G
1HJ8	L	Q	C	L	N	.	.	I	P	I	L	S	Y	S	.	D	C	N	N	S	Y	P	G	M	I	T	N	A	M	.	F	C	A	.	G	Y	.	L	E	G	G	K	D	.	.	.	.	S	C	Q	G	D	S	G	G	P	V	V	C	N	G
TVS4041	V	D	F	M	E	A	R	V	K	Y	V	N	P	.	T	D	C	N	V	W	A	N	F	.	T	T	S	D	K	Q	V	C	S	S	G	Y	S	F	D	S	S	.	D	L	V	T	A	T	C	Q	G	D	S	G	G	P	L	V	W	N	G

	Cβ5							Specificity pocket							Loop2												Cβ6						Cβ6-Cα2				C-terminal α-helix (Cα2)																							
	204	209	210		211	212	213	214	215	216		217	219	220	221A	221B	222		223		224				225	226			227	228	229	230	231	232	233	234	235	236	237	238		239		240					241			242		243					244	245
1BTP	K	L	Q	.	G	I	V	S	W	G	.	S	G	C	A	Q	K	.	N		K	.	.	.	P	G	.	.	V	Y	T	K	V	C	N	Y	V	S	W	I	.	K	.	Q	.	.	.	.	T	.	.	I	.	A	.	.	.	.	S	N
1DPO	E	L	Q	.	G	I	V	S	W	G	.	Y	G	C	A	L	P		.		D	.	N	.	P	G	.	.	V	Y	T	K	V	C	N	Y	V	D	W	I	.	.	.	Q	.	.	D	.	T	.	.	I	A	A	.	.	.	.	.	N
1A0J	Q	L	Q	.	G	V	V	S	W	G	.	Y	G	C	A	Q	R	.	N		K	.	.	.	P	G	.	.	V	Y	T	K	V	C	N	Y	R	S	W	I	.	.	.	.	.	.	.	.	S	S	T	M	.	.	.	.	.	S	S	N
1HJ8	E	L	Q	.	G	V	V	S	W	G	.	Y	G	C	A	E	P	.		.	.	G	N	.	P	G	.	.	V	Y	A	K	V	C	I	F	N	D	W		L.	.	.	.	.	.	.	.	T	S	T	M	.	A	.	.	.	.	S	.
TVS4041	T	.	Q	I	G	I	V	S	F	G	P	S	V	C	G	Q	T	I	V	A	T	G	T	L	P	A	Q	S	V	F	T	D	V	S	H	Y	K	D	W	I	L	K	A	Q	R	G	E	V	T	S	T	I	T	A	T	T	S	S	S	S

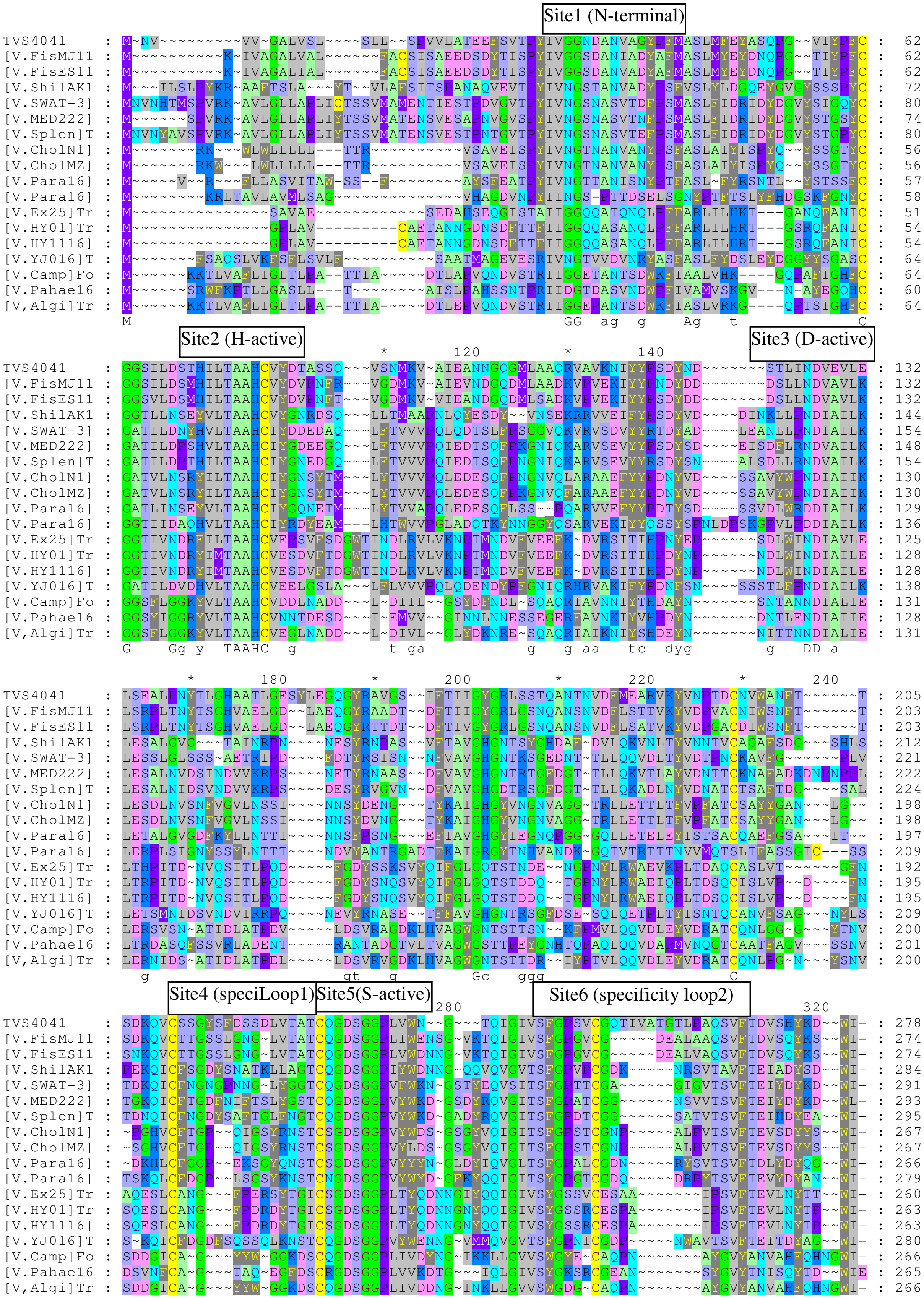
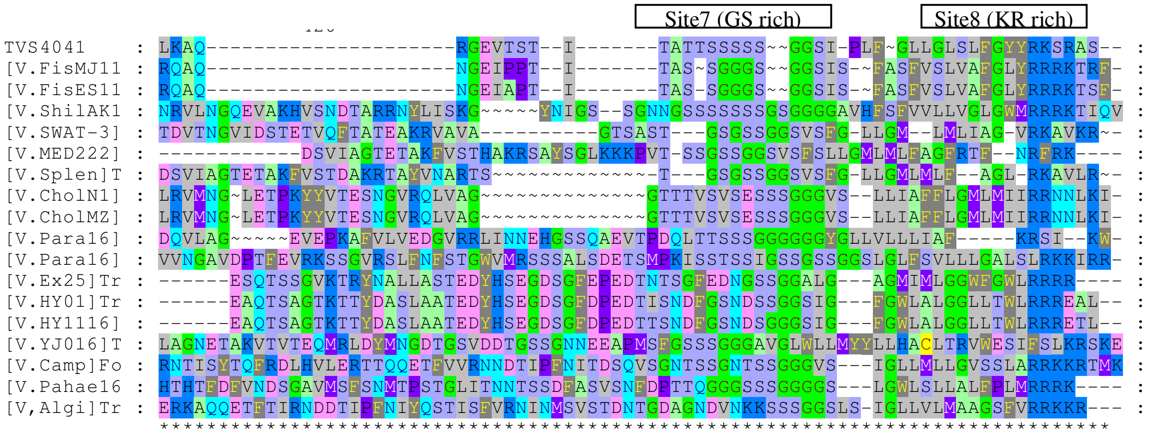


Figure 1.4.2.1.A: alignment with similar species homologs



```
                                                        Site7 (GS rich)         Site8 (KR rich)
                      120
TVS4041     : LKAQ---------------------RGEVTST--I-------TATTSSSSS~~GGSI-PLF~GLLGLSLFGYYRKSRAS-- :
[V.FisMJ11  : RQAQ---------------------NGEIPPT--I-------TAS~SGGGS~~GGSIS-~FASFVSLVAFGLYRRRKTRF- :
[V.FisES11  : RQAQ---------------------NGEIAPT--I-------TAS-SGGGS~~GGSIS-~FASFVSLVAFGLYRRRKTSF- :
[V.ShilAK1  : NRVLNGQEVAKHVSNDTARRNYLISKG~~~~~YNIGS--SGNNGSSSSSSSGSGGGGAVHFSFVVLLVGLGWMRRRKTIQV :
[V.SWAT-3]  : TDVTNGVIDSTETVQFTATEAKRVAVA----------GTSAST---GSGSSGGSVSFG-LLGM--LMLIAG-VRKAVKR~- :
[V.MED222]  : ------------DSVIAGTETAKFVSTHAKRSAYSGLKKKPVT-SSGSSGGSVSFSLLGMLMLFAGFRTF--NRFRK---- :
[V.Splen]T  : DSVIAGTETAKFVSTDAKRTAYVNARTS~~~~~~~~~~~~~~T---GSGSSGGSVSFG-LLGMLMLF--AGL-RKAVLR~- :
[V.CholN1]  : LRVMNG~LETPKYYVTESNGVRQLVAG~~~~~~~~~~~~~~GTTTVSVSESSSGGGVS--LLIAFFLGMLMIIRRNNLKI- :
[V.CholMZ]  : LRVMNG~LETPKYYVTESNGVRQLVAG~~~~~~~~~~~~~~GTTTVSVSESSSGGGVS--LLIAFFLGMLMIIRRNNLKI- :
[V.Para16]  : DQVLAG~~~~~EVEPKAFVLVEDGVRRLINNEHGSSQAEVTPDQLTTSSSGGGGGGYGLLVLLLIAF-----KRSI--KW- :
[V.Para16]  : VVNGAVDPTFEVRKSSGVRSLFNFSTGWVMRSSSALSDETSMPKISSTSSIGSSGSSGGSLGLFSVLLLGALSLRKKIRR- :
[V.Ex25]Tr  : ------ESQTSSGVKTRYNALLASTEDYHSEGDSGFEPEDTNTSGFEDNGSSGGALG---AGMIMLGGWFGWLRRRR---- :
[V.HY01]Tr  : ------EAQTSAGTKTTYDASLAATEDYHSEGDSGFDPEDTISNDFGSNDSSGGSIG---FGWLALGGLLTWLRRREAL-- :
[V.HY1116]  : ------EAQTSAGTKTTYDASLAATEDYHSEGDSGFDPEDTTSNDFGSNDSGGGSIG---FGWLALGGLLTWLRRRETL-- :
[V.YJ016]T  : LAGNETAKVTVTEQMRLDYMNGDTGSVDDTGSSGNNEEAPMSFGSSSGGGAVGLWLLMYYLLHACLTRVWESIFSLKRSKE :
[V.Camp]Fo  : RNTISYTQFRDLHVLERTTQQETFVVRNNDTIPFNITDSQVSGNTSSGNTSSGGGVS---IGLLMLLGVSSLARRKKRTMK :
[V.Pahae16  : HTHTFDFVNDSGAVMSFSNMTPSTGLITNNTSSDFASVSNFDPTTQGGGSSSGGGGS---LGWLSLLALFPLMRRRK---- :
[V,Algi]Tr  : ERKAQQETFTIRNDDTIPFNIYQSTISFVRNINMSVSTDNTGDAGNDVNKKSSSGGSLS-IGLLVLMAAGSFVRRKKR--- :
              *******************************************************************************
```

Figure 1.4.2.2.B: alignment with similar species homologs

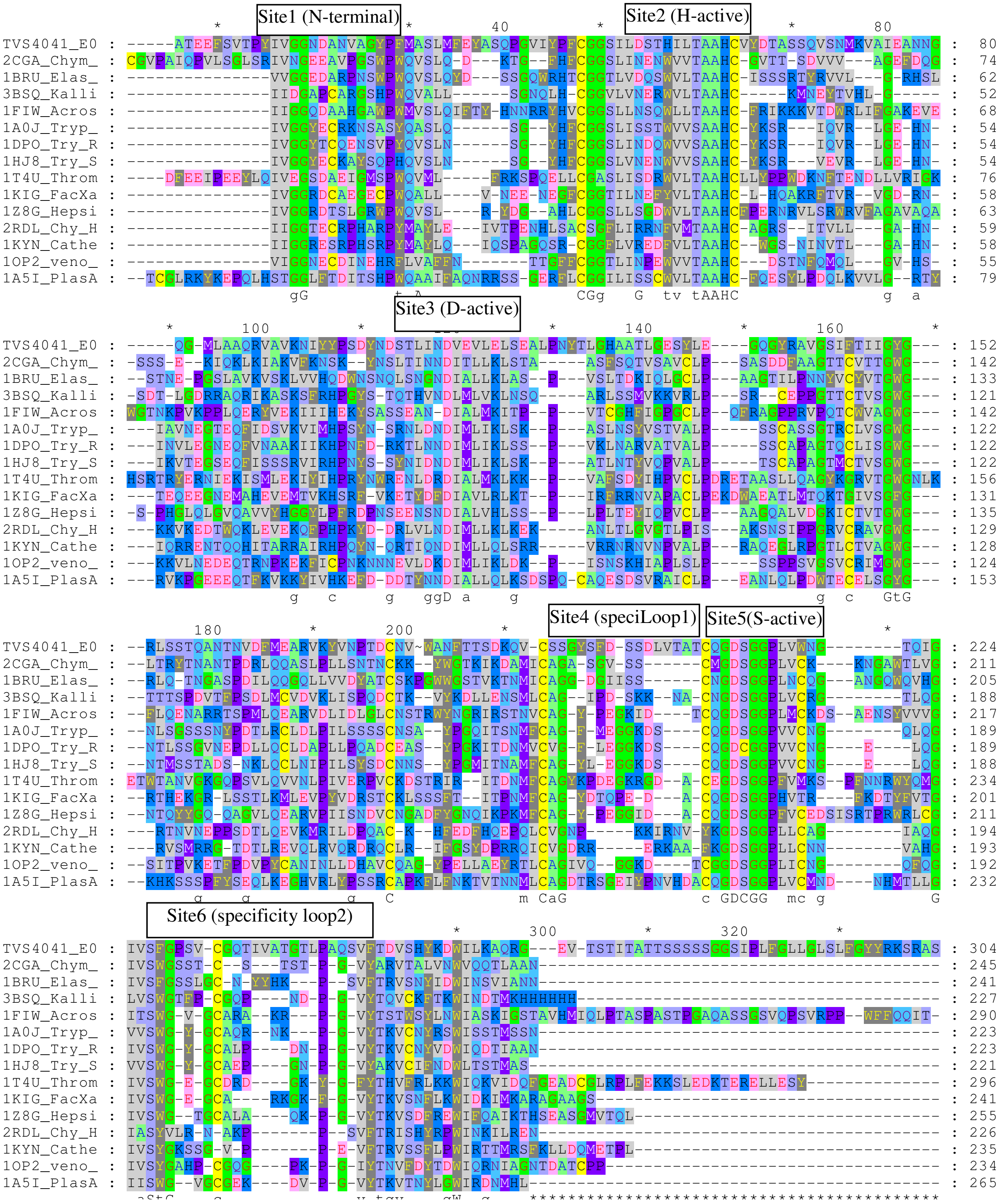
The aligned, sequences were selected from the non redundant blast (http://blast.ncbi.nlm.nih.gov/Blast.cgi), on the bases of their unique similarities from TVS4041, and to see the pattern of evolutionary modification in *Vibrio species*. Entries of sequences from first position are, secreted serine protease [*A. salmonicida*] (323 a.a), elastase 2 [*Vibrio fischeri MJ11*] (319 a.a.) , elastase 2 precursor [*Vibrio fischeri ES114*] (319 a.a), Secreted trypsin-like serine protease [*Vibrio shilonii AK1*] (358 a.a), Trypsin-Like domain [*Vibrionales Bacterium SWAT-3*] (353 a.a), Trypsin-Like domain [*Vibrio MED222*] (355 a.a), Trypsin Like domain [*Vibrio Splendidus 12B01*] (353 a.a), Serine Protease [*Vibrio Cholerae N16961*] (330 a.a), Serine Protease [*Vibrio Cholerae MZO-3*] (330), Serine Protease [*Vibrio Parahaemolyticus16*](334 a.a), chymotrypsin/Hap [*Vibrio Parahaemolyticus16*] (532 a.a), Trypsin Domain [*Vibrio Ex25*] (363 a.a), Trypsin Domain [*Vibrio Harveyi HY01*] (554 a.a), Elastase [*Vibrio Harveyi 1116*] (333 a.a), Secreted Trypsin [*Vibrio Vulnificus YJ016*] (364 a.a), Formyl tetra hydrofolate Deformylase [*Vibrio Campbellii AND4*] (409 a.a), Elastase2 [*Vibrio Parahaemolyticus 16*] (361a.a), Trypsin-Like domain [*Vibrio Alginolyticus 12G01*] (539 a.a).

In overall alignment of active residues HDS (site2, 3 & 5) are well conserved among all entries. The active His in site 2 is much conserved with initial 10 entries in comparison to last 7 entries. The active D (site 3) is conserved in all, except the right hand site residues in TVS4041 are replaced with acidic residue (E) instead of conserved A, K and R residues. Presences of more acidic residues are similar to entries 12 to 18. All six conserved Cys are well conserved in all entries, except in 11th entry the 3rd Cys in chymotrypsins from *Vibrio Parahaemolyticus16*. The first two entries from *Vibrio fischeri* aligned very well with TVS4041, in conserved as well non conserved loop regions, provided they consist of similar number of amino acid; 323 from *A. salmonicida* verses 319 a.a for *V. fischeri*. The identity of TVS4041 and elastase 2 [*Vibrio fischeri MJ11*] is 61% (185/303) and homology is 73% (224/303). The identity of TVS4041 and elastase 2 precursors [*Vibrio fischeri ES114*] is 60% (184/305) and homology is 73% (224/305). The major differences are in both specificity pocket forming loop regions occurs in site 4 and site 6.

The specificity loop1 forming sequence is much similar in length and sequence except the alteration of two Asp D (negatively charged residue) in TVS4041 that are replaced by Gly (neutral residue), in the elastase from *V. fischeri.* As known for elastase activity, the *β*-branched residues like Val and Thr occupies the specificity pocket, hence do not leave the space for any bulky, or charged residue to be able to fit into specificity pocket (Branden and Tooze 1999). In case of TVS4041 "LVTAT" sequence before conserved Cys is the same as in sequences from *V. fischeri.* But the presence of two Asp (D) in TVS4041 makes it different from elastases of *V. fischeri*, but similar to entries 13, 14, 15, 17. It can be suspected that these Asp will create the electronegative charge inside the specificity pocket hence this enzyme will act more like trypsin, having attraction to positively charge residue, to be fixed in its specificity pocket (Arg/Lys as P1).

Second specificity loop is much more unique compared to all closer species since it has long insertion that do not resemble with any of the closer species sequence. Interestingly it can be noted that insertion contains T, I, V and A residues, that are similar in contents with the loop1 insertions. Hence it can be imagined that this specificity pocket will be filled with T, V, A, L and I. Furthermore, the insertions in both specificity pockets will increase the size of this specificity pocket, and relative appearance of these residues in the brim of specificity pocket will determine the P1 affinity.

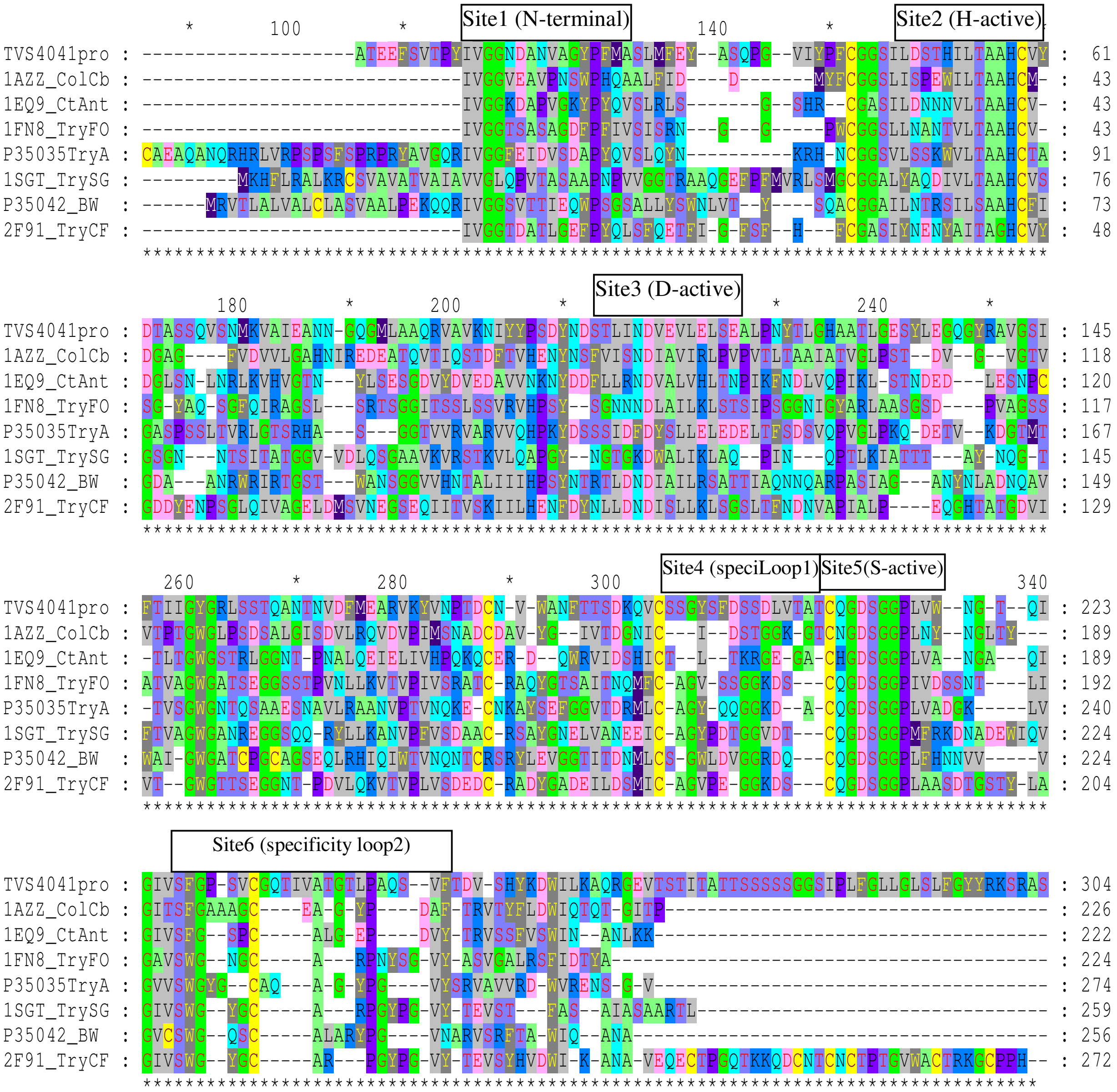
The interesting feature of the closer homologus species alignment is their *C*-terminal sequence. Irrespective of their varying length, 319 in elastase (*Vibrio fischeri ES114)* to 554 in elastase (*Vibrio Harveyi HY01),* the GS rich site 7 and RK rich site 8 and hydrophobic sequences between them are conserved in the *C*-terminal of these species, seems to be specially modified by nature in special purpose of function. When this pentameric serine followed by two consequent glycine and then serine from TVS4041 was blast searched into NCBI data base many *Drosophila spp.* sequences were found to contain 100% similar region. From vertebral species this kind of multi serine and glycine region was also observed in extended *C*-terminal of ram spermatozoa (figure: 4.4.2.3). This represents the ancient evolutionary relation of these kind of sequences for special purpose like target attachment, secretion or temporary inhibition.



```
                          Site1 (N-terminal)                    Site2 (H-active)
                      *                   *        40         *                   *        80
TVS4041_E0 : -----ATEEFSVTPYIVGGNDANVAGYPFMASLMFEYASQPGVIYPFCGGSILDSTHILTAAHCVYDTASSQVSNMKVAIEANNG :  80
2CGA_Chym_ : CGVPAIQPVLSGLSRIVNGEEAVPGSWPWQVSLQ-D---KTG--FHFCGGSLINENWVVTAAHC-GVTT-SDVVV---AGEFDQG :  74
1BRU_Elas_ : ---------------VVGGEDARPNSWPWQVSLQYD---SSGQWRHTCGGTLVDQSWVLTAAHC-ISSSRTYRVVL---G-RHSL :  62
3BSQ_Kalli : ---------------IIDGAPCARGSHPWQVALL------SGNQLH-CGGVLVNERWVLTAAHC-----KMNEYTVHL-G----- :  52
1FIW_Acros : ---------------IIGGQDAAHGAWPWMVSLQIFTY-HNNRRYHVCGGSLLNSQWLLTAAHC-FRIKKKVTDWRLIFGAKEVE :  68
1A0J_Tryp_ : ---------------IVGGYECRKNSASYQASLQ------SG--YHFCGGSLISSTWVVSAAHC-YKSR---IQVR--LGE-HN- :  54
1DPO_Try_R : ---------------IVGGYTCQENSVPYQVSLN------SG--YHFCGGSLINDQWVVSAAHC-YKSR---IQVR--LGE-HN- :  54
1HJ8_Try_S : ---------------IVGGYECKAYSQPHQVSLN------SG--YHFCGGSLVNENWVVSAAHC-YKSR---VEVR--LGE-HN- :  54
1T4U_Throm : ----DFEEIPEEYLQIVEGSDAEIGMSPWQVML-----FRKSPQELLCGASLISDRWVLTAAHCLLYPPWDKNFTENDLLVRIGK :  76
1KIG_FacXa : ---------------IVGGRDCAEGECPWQALL----V-NEE-NEGFCGGTILNEFYVLTAAHC-L-HQAKRFTVR--VGD-RN- :  58
1Z8G_Hepsi : ---------------IVGGRDTSLGRWPWQVSL----R-YDG--AHLCGGSLLSGDWVLTAAHCFPERNRVLSRWRVFAGAVAQA :  63
2RDL_Chy_H : ---------------IIGGTECRPHARPYMAYLE---IVTPENHLSACSGFLIRRNFVMTAAHC-AGRS--ITVLL---GA-HN- :  59
1KYN_Cathe : ---------------IIGGRESRPHSRPYMAYLQ---IQSPAGQSR-CGGFLVREDFVLTAAHC--WGS-NINVTL---GA-HN- :  58
1OP2_veno_ : ---------------VIGGNECDINEHRFLVAFFN-------TTGFFCGGTLINPEWVVTAAHC---DSTNFQMQL---GV-HS- :  55
1A5I_PlasA : --TCGLRKYKEPQLHSTGGLFTDITSHPWQAAIFAQNRRSS-GERFLCGGILISSCWVLTAAHC-FQESYLPDQLKVVLG--RTY :  79
                              gG         + A                CGg   G tv  tAAHC               g  a

                                         Site3 (D-active)
                 *       100         *       120         *       140         *       160         *
TVS4041_E0 : -----QG-MLAAQRVAVKNIYYPSDYNDSTLINDVEVLELSEALPNYTLGHAATLGESYLE----GQGYRAVGSIFTIIGYG--- : 152
2CGA_Chym_ : -SSS-E--KIQKLKIAKVFKNSK--YNSLTINNDITLLKLSTA-----ASFSQTVSAVCLP---SASDDFAAGTTCVTTGWG--- : 142
1BRU_Elas_ : --STNE-PGSLAVKVSKLVVHQDWNSNQLSNGNDIALLKLAS--P---VSLTDKIQLGCLP---AAGTILPNNYVCYVTGWG--- : 133
3BSQ_Kalli : -SDT-LGDRRAQRIKASKSFRHPGYS-TQTHVNDLMLVKLNSQ-----ARLSSMVKKVRLP---SR--CEPPGTTCTVSGWG--- : 121
1FIW_Acros : WGTNKPVKPPLQERYVEKIIIHEKYSASSEAN-DIALMKITP--P---VTCGHFIGPGCLP--QFRAGPPRVPQTCWVAGWG--- : 142
1A0J_Tryp_ : ---IAVNEGTEQFIDSVKVIMHPSYN-SRNLDNDIMLIKLSK--P---ASLNSYVSTVALP-----SSCASSGTRCLVSGWG--- : 122
1DPO_Try_R : ---INVLEGNEQFVNAAKIIKHPNFD-RKTLNNDIMLIKLSS--P---VKLNARVATVALP-----SSCAPAGTQCLISGWG--- : 122
1HJ8_Try_S : ---IKVTEGSEQFISSSRVIRHPNYS-SYNIDNDIMLIKLSK--P---ATLNTYVQPVALP-----TSCAPAGTMCTVSGWG--- : 122
1T4U_Throm : HSRTRYERNIEKISMLEKIYIHPRYNWRENLDRDIALMKLKK--P---VAFSDYIHPVCLPDRETAASLLQAGYKGRVTGWGNLK : 156
1KIG_FacXa : ---TEQEEGNEMAHEVEMTVKHSRF-VKETYDFDIAVLRLKT--P---IRFRRNVAPACLPEKDWAEATLMTQKTGIVSGFG--- : 131
1Z8G_Hepsi : -S-PHGLQLGVQAVVYHGGYLPFRDPNSEENSNDIALVHLSS--P---LPLTEYIQPVCLP---AAGQALVDGKICTVTGWG--- : 135
2RDL_Chy_H : ---KKVKEDTWQKLEVEKQFPHPKYD-DRLVLNDIMLLKLKEK-----ANLTLGVGTLPIS---AKSNSIPPGRVCRAVGWG--- : 129
1KYN_Cathe : ---IQRRENTQQHITARRAIRHPQYN-QRTIQNDIMLLQLSRR-----VRRNRNVNPVALP---RAQEGLRPGTLCTVAGWG--- : 128
1OP2_veno_ : ---KKVLNEDEQTRNPKEKFICPNKNNNEVLDKDIMLIKLDK--P---ISNSKHIAPLSLP-----SSPPSVGSVCRIMGWG--- : 124
1A5I_PlasA : ---RVKPGEEEQTFKVKKYIVHKEFD-DDTYNNDIALLQLKSDSPQ-CAQESDSVRAICLP---EANLQLPDWTECELSGYG--- : 153
                              g   c     g   ggD a    g                               g  c   GtG

                                                        Site4 (speciLoop1) Site5(S-active)
                    180         *       200         *                                       *
TVS4041_E0 : --RLSSTQANTNVDFMEARVKYVNPTDCNV~WANFTTSDKQV-CSSGYSFD-SSDLVTATCQGDSGGPLVWNG--------TQIG : 224
2CGA_Chym_ : --LTRYTNANTPDRLQQASLPLLSNTNCKK--YWGTKIKDAMICAGA-SGV-SS------CMGDSGGPLVCK----KNGAWTLVG : 211
1BRU_Elas_ : --RLQ-TNGASPDILQQGQLLVVDYATCSKPGWWGSTVKTNMICAGG-DGIISS------CNGDSGGPLNCQG---ANGQWQVHG : 205
3BSQ_Kalli : --TTTSPDVTFPSDLMCVDVKLISPQDCTK--VYKDLLENSMLCAG--IPD-SKK--NA-CNGDSGGPLVCRG--------TLQG : 188
1FIW_Acros : --FLQENARRTSPMLQEARVDLIDLGLCNSTRWYNGRIRSTNVCAG-Y-PEGKID----TCQGDSGGPLMCKDS--AENSYVVVG : 217
1A0J_Tryp_ : --NLSGSSSNYPDTLRCLDLPILSSSSCNSA--YPGQITSNMFCAG-F-MEGGKDS----CQGDSGGPVVCNG--------QLQG : 189
1DPO_Try_R : --NTLSSGVNEPDLLQCLDAPLLPQADCEAS--YPGKITDNMVCVG-F-LEGGKDS----CQGDCGGPVVCNG----E----LQG : 189
1HJ8_Try_S : --NTMSSTADS-NKLQCLNIPILSYSDCNNS--YPGMITNAMFCAG-YL-EGGKDS----CQGDSGGPVVCNG----E----LQG : 188
1T4U_Throm : ETWTANVGKGQPSVLQVVNLPIVERPVCKDSTRIR--ITDNMFCAGYKPDEGKRGD--A-CEGDSGGPFVMKS--PFNNRWYQMG : 234
1KIG_FacXa : --RTHEKGR-LSSTLKMLEVPYVDRSTCKLSSSFT--ITPNMFCAG-YDTQPE-D---A-CQGDSGGPHVTR----FKDTYFVTG : 201
1Z8G_Hepsi : --NTQYYGQ-QAGVLQEARVPIISNDVCNGADFYGNQIKPKMFCAG-Y-PEGGID---A-CQGDSGGPFVCEDSISRTPRWRLCG : 211
2RDL_Chy_H : ---RTNVNEPPSDTLQEVKMRILDPQAC-K--HFEDFHQEPQLCVGNP-----KKIRNV-YKGDSGGPLLCAG--------IAQG : 194
1KYN_Cathe : ---RVSMRRG-TDTLREVQLRVQRDRQCLR--IFGSYDPRRQICVGDRR-----ERKAA-FKGDSGGPLLCNN--------VAHG : 193
1OP2_veno_ : --SITPVKETFPDVPYCANINLLDHAVCQAG-YPELLAEYRTLCAGIVQ--GGKD----TCGGDSGGPLICNG--------QFQG : 192
1A5I_PlasA : --KHKSSSPFYSEQLKEGHVRLYPSSRCAPKFLFNKTVTNNMLCAGDTRSGEIYPNVHDACQGDSGGPLVCMND----NHMTLLG : 232
                    g    g          g   C             m CaG              c GDCGG  mc g           G

               Site6 (specificity loop2)
                                              *       300         *       320         *
TVS4041_E0 : IVSFGPSV-CGQTIVATGTLPAQSVFTDVSHYKDWILKAQRG---EV-TSTITATTSSSSSGGSIPLFGLLGLSLFGYYRKSRAS : 304
2CGA_Chym_ : IVSWGSST-C--S---TST-P-G-VYARVTALVNWVQQTLAAN------------------------------------------ : 245
1BRU_Elas_ : IVSFGSSLGC-N-YYHK---P--SVFTRVSNYIDWINSVIANN------------------------------------------ : 241
3BSQ_Kalli : LVSWGTFP-CGQP----ND-P-G-VYTQVCKFTKWINDTMKHHHHHH-------------------------------------- : 227
1FIW_Acros : ITSWG-V-GCARA--KR---P-G-VYTSTWSYLNWIASKIGSTAVHMIQLPTASPASTPGAQASSGSVQPSVRPP--WFFQQIT- : 290
1A0J_Tryp_ : VVSWG-Y-GCAQR--NK---P-G-VYTKVCNYRSWISSTMSSN------------------------------------------ : 223
1DPO_Try_R : IVSWG-Y-GCALP----DN-P-G-VYTKVCNYVDWIQDTIAAN------------------------------------------ : 223
1HJ8_Try_S : VVSWG-Y-GCAEP----GN-P-G-VYAKVCIFNDWLTSTMAS------------------------------------------- : 221
1T4U_Throm : IVSWG-E-GCDRD----GK-Y-G-FYTHVFRLKKWIQKVIDQFGEADCGLRPLFEKKSLEDKTERELLESY-------------- : 296
1KIG_FacXa : IVSWG-E-GCA----RKGK-F-G-VYTKVSNFLKWIDKIMKARAGAAGS------------------------------------ : 241
1Z8G_Hepsi : IVSWG--TGCALA----QK-P-G-VYTKVSDFREWIFQAIKTHSEASGMVTQL-------------------------------- : 255
2RDL_Chy_H : IASYVLR-N-AKP-------P--SVFTRISHYRPWINKILREN------------------------------------------ : 226
1KYN_Cathe : IVSYGKSSG-V-P-------P-E-VFTRVSSFLPWIRTTMRSFKLLDQMETPL-------------------------------- : 235
1OP2_veno_ : IVSYGAHP-CGQG----PK-P-G-IYTNVFDYTDWIQRNIAGNTDATCPP----------------------------------- : 234
1A5I_PlasA : IISWG--VGCGEK----DV-P-G-VYTKVTNYLGWIRDNMHL------------------------------------------- : 265
              aStG    c              v tgv    gW  g    *******************************************
```

Figure 4.4.2.3: alignment of TVS4041 with homologous serine proteases form vertebral source

2CGA is Chymotrypsin from Bovine, 1FIWβ is Acrosin from Ram Spermatozoa, 1BRU is Elastase from Porcine, 1DPO is Trypsin from Rat, 1A0J is Trypsin from ColdFish, 1HJ8 is Trypsin from Atlantic Salmon, 1T4UT is Thrombin from Humans, 1A5I is PlasActivator from Bat, 1KIG is FacXa from Bov, 3BSQ is Kallikrein from Human, 2RDL is Chymotrypsin from Hamster, 1Z8G Hepsin, 1KYN is Cathepsin G from human, 1OP2 is venom from Snake.



```
                                           Site1 (N-terminal)                      Site2 (H-active)
                 *       100         *                           140         *
TVS4041pro : --------------------ATEEFSVTPYIVGGNDANVAGYPFMASLMFEY--ASQPG--VIYPFCGGSILDSTHILTAAHCVY :  61
1AZZ_ColCb : ------------------------------IVGGVEAVPNSWPHQAALFID----D-------MYFCGGSLISPEWILTAAHCM- :  43
1EQ9_CtAnt : ------------------------------IVGGKDAPVGKYPYQVSLRLS-------G--SHR--CGASILDNNNVLTAAHCV- :  43
1FN8_TryFO : ------------------------------IVGGTSASAGDFPFIVSISRN---G---G-----PWCGGSLLNANTVLTAAHCV- :  43
P35035TryA : CAEAQANQRHRLVRPSPSFSPRPRYAVGQRIVGGFEIDVSDAPYQVSLQYN----------KRH-NCGGSVLSSKWVLTAAHCTA :  91
1SGT_TrySG : ---------MKHFLRALKRCSVAVATVAIAVVGLQPVTASAAPNPVVGGTRAAQGEFPFMVRLSMGCGGALYAQDIVLTAAHCVS :  76
P35042_BW  : ------MRVTLALVALCLASVAALPEKQQRIVGGSVTTIEQWPSGSALLYSWNLVT--Y----SQACGGAILNTRSILSAAHCFI :  73
2F91_TryCF : ------------------------------IVGGTDATLGEFPYQLSFQETFI-G-FSF--H---FCGASIYNENYAITAGHCVY :  48
             *************************************************************************************

                                                       Site3 (D-active)
                    180         *       200         *                   *       240         *
TVS4041pro : DTASSQVSNMKVAIEANN-GQGMLAAQRVAVKNIYYPSDYNDSTLINDVEVLELSEALPNYTLGHAATLGESYLEGQGYRAVGSI : 145
1AZZ_ColCb : DGAG----FVDVVLGAHNIREDEATQVTIQSTDFTVHENYNSFVISNDIAVIRLPVPVTLTAAIATVGLPST--DV--G--VGTV : 118
1EQ9_CtAnt : DGLSN-LNRLKVHVGTN---YLSESGDVYDVEDAVVNKNYDDFLLRNDVALVHLTNPIKFNDLVQPIKL-STNDED---LESNPC : 120
1FN8_TryFO : SG-YAQ-SGFQIRAGSL---SRTSGGITSSLSSVRVHPSY--SGNNNDLAILKLSTSIPSGGNIGYARLAASGSD----PVAGSS : 117
P35035TryA : GASPSSLTVRLGTSRHA---S---GGTVVRVARVVQHPKYDSSSIDFDYSLLELEDELTFSDSVQPVGLPKQ-DETV--KDGTMT : 167
1SGT_TrySG : GSGN---NTSITATGGV-VDLQSGAAVKVRSTKVLQAPGY--NGTGKDWALIKLAQ--PIN---QPTLKIATTT---AY-NQG-T : 145
P35042_BW  : GDA---ANRWRIRTGST---WANSGGVVHNTALIIIHPSYNTRTLDNDIAILRSATTIAQNNQARPASIAG---ANYNLADNQAV : 149
2F91_TryCF : GDDYENPSGLQIVAGELDMSVNEGSEQIITVSKIILHENFDYNLLDNDISLLKLSGSLTFNDNVAPIALP----EQGHTATGDVI : 129
             *************************************************************************************

                                                              Site4 (speciLoop1)Site5(S-active)
               260         *       280         *       300                                     340
TVS4041pro : FTIIGYGRLSSTQANTNVDFMEARVKYVNPTDCN-V-WANFTTSDKQVCSSGYSFDSSDLVTATCQGDSGGPLVW--NG-T--QI : 223
1AZZ_ColCb : VTPTGWGLPSDSALGISDVLRQVDVPIMSNADCDAV-YG--IVTDGNIC---I--DSTGGK-GTCNGDSGGPLNY--NGLTY--- : 189
1EQ9_CtAnt : -TLTGWGSTRLGGNT-PNALQEIELIVHPQKQCER-D--QWRVIDSHICT--L--TKRGE-GA-CHGDSGGPLVA--NGA---QI : 189
1FN8_TryFO : ATVAGWGATSEGGSSTPVNLLKVTVPIVSRATC-RAQYGTSAITNQMFC-AGV-SSGGKDS---CQGDSGGPIVDSSNT----LI : 192
P35035TryA : -TVSGWGNTQSAAESNAVLRAANVPTVNQKE-CNKAYSEFGGVTDRMLC-AGY-QQGGKD--A-CQGDSGGPLVADGK-----LV : 240
1SGT_TrySG : FTVAGWGANREGGSQQ-RYLLKANVPFVSDAAC-RSAYGNELVANEEIC-AGYPDTGGVDT---CQGDSGGPMFRKDNADEWIQV : 224
P35042_BW  : WAI-GWGATCPGCAGSEQLRHIQIWTVNQNTCRSRYLEVGGTITDNMLCS-GWLDVGGRDQ---CQGDSGGPLFHNNVV-----V : 224
2F91_TryCF : VT--GWGTTSEGGNT-PDVLQKVTVPLVSDEDC-RADYGADEILDSMIC-AGVPE-GGKDS---CQGDSGGPLAASDTGSTY-LA : 204
             *************************************************************************************

                Site6 (specificity loop2)
TVS4041pro : GIVSFGP-SVCGQTIVATGTLPAQS--VFTDV-SHYKDWILKAQRGEVTSTITATTSSSSSGGSIPLFGLLGLSLFGYYRKSRAS : 304
1AZZ_ColCb : GITSFGAAAGC----EA-G-YP----DAF-TRVTYFLDWIQTQT-GITP------------------------------------ : 226
1EQ9_CtAnt : GIVSFG--SPC-----ALG-EP----DVY-TRVSSFVSWIN--ANLKK------------------------------------- : 222
1FN8_TryFO : GAVSWG--NGC-----A---RPNYSG-VY-ASVGALRSFIDTYA----------------------------------------- : 224
P35035TryA : GVVSWGYG--CAQ---A-G-YPG----VYSRVAVVRD-WVRENS-G-V------------------------------------- : 274
1SGT_TrySG : GIVSWG--YGC-----A---RPGYPG-VY-TEVST---FAS--AIASAARTL--------------------------------- : 259
P35042_BW  : GVCSWG--QSC-----ALARYPG----VNARVSRFTA-WIQ--ANA--------------------------------------- : 256
2F91_TryCF : GIVSWG--YGC-----AR---PGYPG-VY-TEVSYHVDWI-K-ANA-VEQECTPGQTKKQDCNTCNCTPTGVWACTRKGCPPH-- : 272
             ************************************************************************************
```

Figure 1.4.2.4.: alignment with similar sequences from invertebral species

Above aligned, sequences were selected from homologous sequences from invertebral species. Where, TVS4041 is secreted serine protease [*A. salmonicida*] (294 a.a), 1AZZ ColCb is sequence of collagnase [*Crab*] (226 a.a), 1EQ9_CtAnt is sequence of Chymotrypsin from [*Fire Ant*] (222a.a), 1FN8 TrypFO is sequence of trypsin [*Fusarium oxysporium*] (224 a.a)**,** P35035TryAG is sequence of trypsin [*Anopheles gambiae*] (227a.a), 1SGT is sequence of trypsin [*Streptomyces Griseus*] (240a.a), P35042 BW is sequence of trypsin [*bud worm*] (232a.a), 2F91 TryCF is sequence of trypsin [*Cry Fish*] (272a.a).

From the above two alignments, it can be seen that vertebral source serine proteases are, although represented from different type and distant species, but they poses more conservation than invertebral sources serine proteases. Moreover, it can be seen that the number of Cys making disulphide bonds are much grater in vertebral species, compared to the serine proteases from invertebral species. Since disulphide bonds increase the protein rigidity and stability, based on that it can be predicted that serine proteases from vertebral source would be more stable than invertebral source serine proteases.

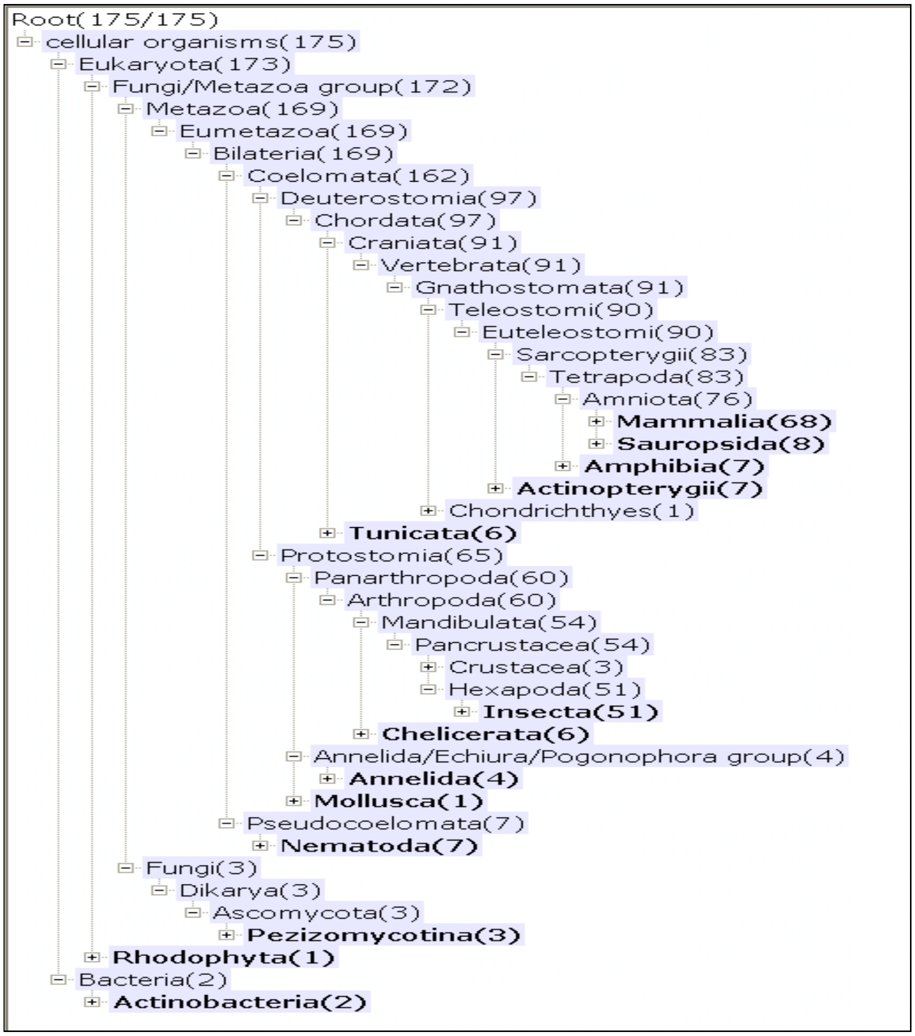
The presence of Pro, are also used to conserve the structural confirmation and stability of any protein structure (Branden and Tooze 1999). Numbers of Pro in vertebral serine proteases are more conserved and grater in number, eight in comparison to four conserved Pro in invertebral species. In contrast to Pro, number of Gly is known for conformational freedom. It can be seen that 23 Gly in TVS4041 are much well aligned to vertebral serine proteases, in comparison to invertebral source serine proteases.

Among all entries of vertebral and invertebral and similar species source serine proteases, the number of residues in specificity loop 1 and specificity loop 2 are most abundant in TVS4041. The number of Ser and Thr are much abundant in these specificity loops. Moreover, from both specificity loops no positively charged residue (KRH) can be seen, as can be seen in other sequences from different types and varying sources. This character is in common with *Bovine Chymotrypsin* (2GCA). In case of invertebral alignment, Ala appears as a conserved residue of specificity loop2 that is also similar in TVS4041 and only for some of the vertebral source serine proteases.

Aligned sequences were evaluated for their evolutionary relations in BioEdit by Neighbor-Joining UPGMA method, version 3.6a2.1 (Hall 1997). From vertebral alignment TVS4041, predicted to be closer to Hamster Chymotrypsin (2RLD) and Human Cathepsin G (1KYN). While, from invertebral sequences, TVS4041 was found closely linked to Crab collagnase (1AZZ) FireAnt chymotrypsin (1EQ9) and bud worm trypsin (P35042).

4.4.3 Phylogenetic relational analysis:

Phylogenic relation of TVS4041 was evaluated by Conserved Domain Database, NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrv.cgi). CDD analysis revels TVS4041 as Try-SPc domain belonging to family CD00190 and superfamily cl00149, mostly reported to be produced as inactive zymogens. The taxonomic tree indicates the presence of such domain throughout the living organisms mostly in class *Mammalia* (68) and *Insecta* (51). These kind of Try-SPc domains have been found to be rare for bacteria, limitated to only *actinobacteria* (2) (figure 4.4.3.A) (Marchler-Bauer, Anderson et al. 2007)**.**



```
Root(175/175)
⊟ cellular organisms(175)
  ⊟ Eukaryota(173)
    ⊟ Fungi/Metazoa group(172)
      ⊟ Metazoa(169)
        ⊟ Eumetazoa(169)
          ⊟ Bilateria(169)
            ⊟ Coelomata(162)
              ⊟ Deuterostomia(97)
                ⊟ Chordata(97)
                  ⊟ Craniata(91)
                    ⊟ Vertebrata(91)
                      ⊟ Gnathostomata(91)
                        ⊟ Teleostomi(90)
                          ⊟ Euteleostomi(90)
                            ⊟ Sarcopterygii(83)
                              ⊟ Tetrapoda(83)
                                ⊟ Amniota(76)
                                  ⊞ Mammalia(68)
                                  ⊞ Sauropsida(8)
                                ⊞ Amphibia(7)
                            ⊞ Actinopterygii(7)
                        ⊞ Chondrichthyes(1)
                  ⊞ Tunicata(6)
              ⊟ Protostomia(65)
                ⊟ Panarthropoda(60)
                  ⊟ Arthropoda(60)
                    ⊟ Mandibulata(54)
                      ⊟ Pancrustacea(54)
                        ⊞ Crustacea(3)
                        ⊟ Hexapoda(51)
                          ⊞ Insecta(51)
                    ⊞ Chelicerata(6)
                ⊟ Annelida/Echiura/Pogonophora group(4)
                  ⊞ Annelida(4)
                ⊞ Mollusca(1)
            ⊟ Pseudocoelomata(7)
              ⊞ Nematoda(7)
        ⊟ Fungi(3)
          ⊟ Dikarya(3)
            ⊟ Ascomycota(3)
              ⊞ Pezizomycotina(3)
    ⊞ Rhodophyta(1)
  ⊟ Bacteria(2)
    ⊞ Actinobacteria(2)
```

Figure 4.4.3.A: Distribution of TVS4041 like enzyme in living organisms.
“Taxonomy tree of cd00190, source:

The phylogenic tree was constructed with 175 homologous sequences of conserve domain family “**cd00190”.** Placement of TVS4041 (Query, Red in color) was in the parents sequences of family. These early evolutionary event of family includes sequences from *C.elegance, M. sexta, H. armigere* and *P. purpurea* (Marchler-Bauer, Anderson et al. 2007)**.** The other sequences indicated with red arrows were also taken in consideration during invertebrel alignment (figure 1.2.4.4)

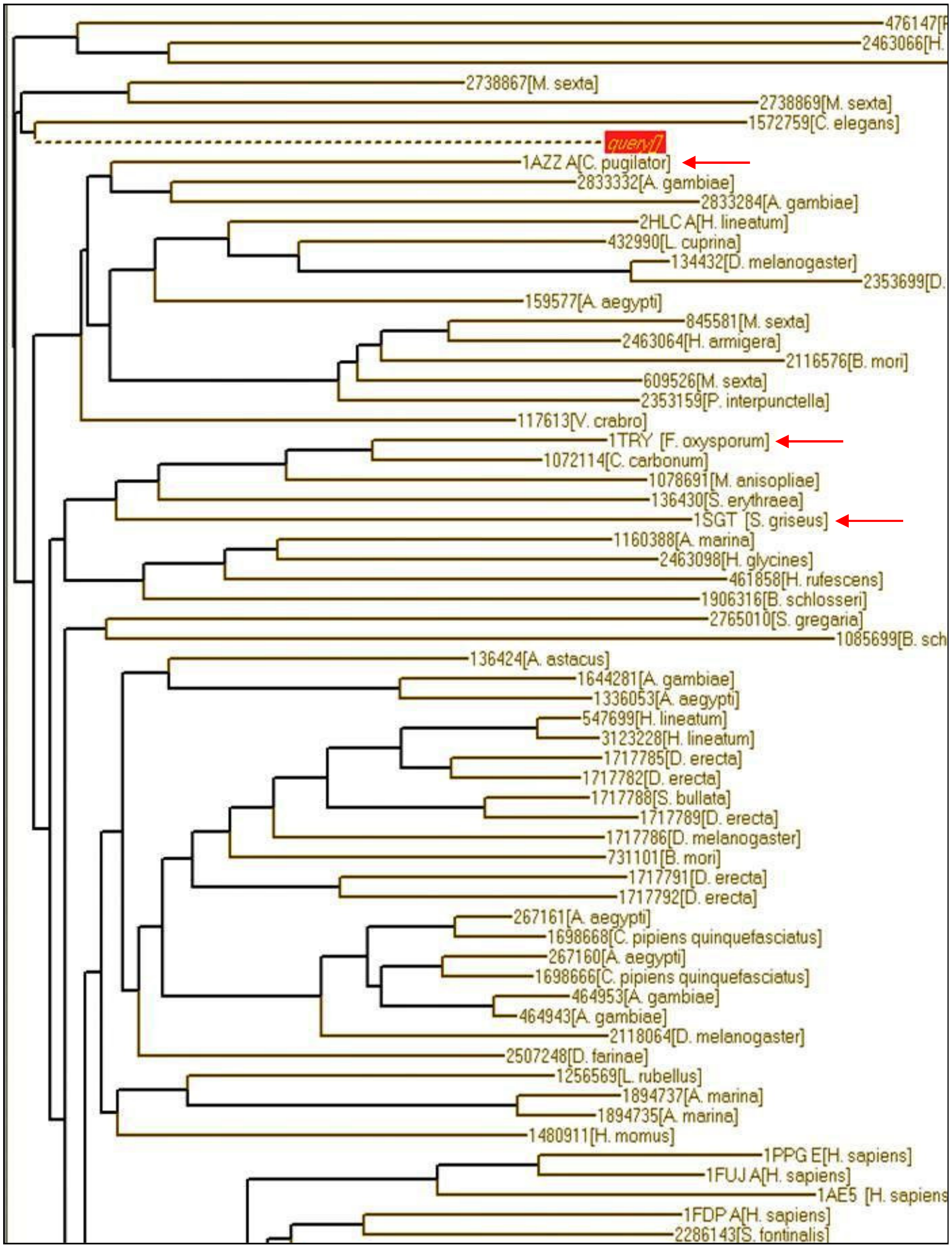


Figure 4.4.3.B: position of TVS4041 (red in color) in Phylogenic tree of cd00190. Source: (Marchler-Bauer, Anderson et al. 2007).

4.4.4.1. Homology-based model building using ICM™ Bioinformatics:

The amino acid sequence of TVS4041 have been blasted toward the Protein Databank (PDB, www.pdb.org) (Altschul, Gish et al. 1990) to identify the closest possible homologs with available structures in the PDB. The closest possible structure was found **1fn8** (Rypniewski, Ostergaard et al. 2001) (trypsin from *fusarium oxysporium*) as 'template' with identities 69/223 (30%), homology = 110/223 (49%). During the first step of model building, the 'pairwise alignment' has been performed using ClustalW (Thompson, Gibson et al. 2002).

Then the secondary structure of the modeled protein (TVS4041-1fn8) has been predicted using the ICM™ Bioinformatics module. Utilizing these predicted secondary structural elements and 1fn8 as template, the final building and the model refinement also have been performed using the 3D space of ICM™ Bioinformatics module. The ICM™ program constructs the molecular model by homology from core sections defined by the average of Cα atom positions in conserved regions (figure: 4.4.4.1 A&B).

The 3D structures of 'TVS4041-1fn8' model and the 1fn8 were superimposed and their RMSD (root mean squared deviation) between backbone Cα atoms value was calculated utilizing DaliLite (Holm and Park 2000) (http://www.ebi.ac.uk/Tools/dalilite/index.html) in order to compare their overall 3D structures and active site architecture. RMSD between 'TVS4041-1fn8' **model** and **1fn8** was found as 0.7 Å and Z-score was found as 40.9 Z-Scores (Holm and Park 2000).

RMSD value of 'TVS4041-1fn8' was below 1.0 and Z- score is above 40, that represents the overall good quality of the predicted model. Although there are some differences in some of the loop regions, but most of the important foldings of the model, like helices, β-sheets were smoothly superimposed on the 1fn8, which is shown in Figure 4.4.4.2, where the red colored structure is the model 'TVS4041-1fn8' and the light blue is the 'template' (1fn8).

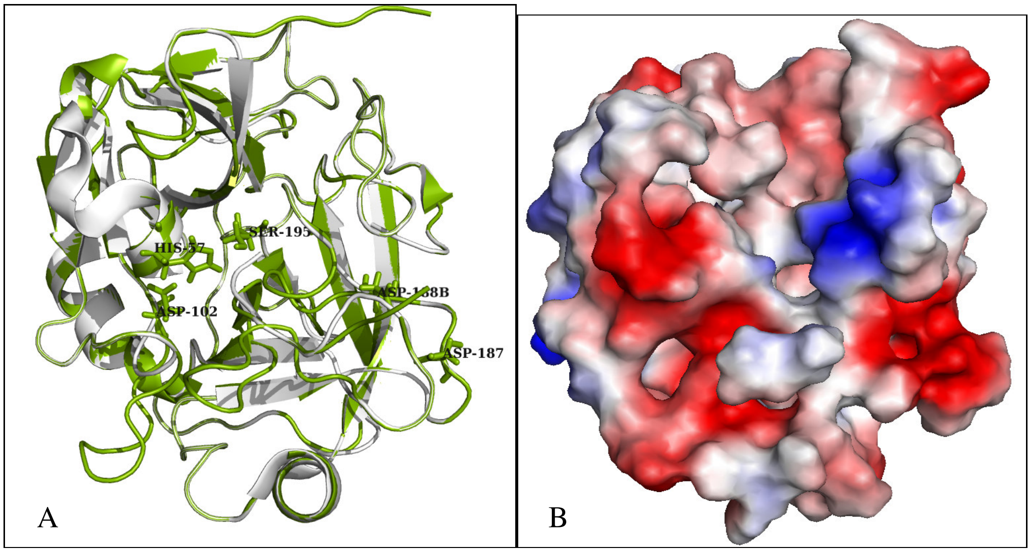


Figure 4.4.4.1: (A): Superimposition of template **1fn8** (gray) and model 'TVS4041-1fn8' (green). (B): surface representation of 'TVS4041-1fn8'.

Beside RMSD and Z-score two more parameters were used to assess the quality of modeled protein. PROCHECK was used to check the quality of modeled protein based on phi, psi and chi 1 torsion angles (Morris, MacArthur et al. 1992). The procheck score for modeled 'TVS4041-1fn8' was calculated as 88.2% and 8.7% in the favoured and allowed regions, where as the value for template (1fn8) was 90.0% and 10% for the same regions. But on the other hand side 'TVS4041-1fn8' also contained the 1.9% and 1.2% in of residues in generously allowed and disallowed region respectivly. That represents the bad sectors in model 'TVS4041-1fn8'.

4.4.4.2. Model building through Swiss-Model workspace:

The 'TVS4041-1fn8' model builded through ICM server was showing significantly good quality, but was broken in substrate specificity pocket. Based on this another model was builded using Swiss-Model server represented in figure 4.4.2 A&B (http://**swissmodel**.expasy.org/workspace/). The automated generated model (TVS4041-1OP8) was made using pdb entry 1OP8 (Hink-Schauer, Estebanez-Perpina et al. 2003), as template this attempt resulted in complete carbon alpha chain and complete loop in specificity pocket. The RMSD value was found 1.4 Å that is comparatively higher than 0.7 Å for ‘TVS4041-1fn8’. The Z-score was found as 50.0 (Holm and Park 2000), that seems better than TVS4041-1fn8.

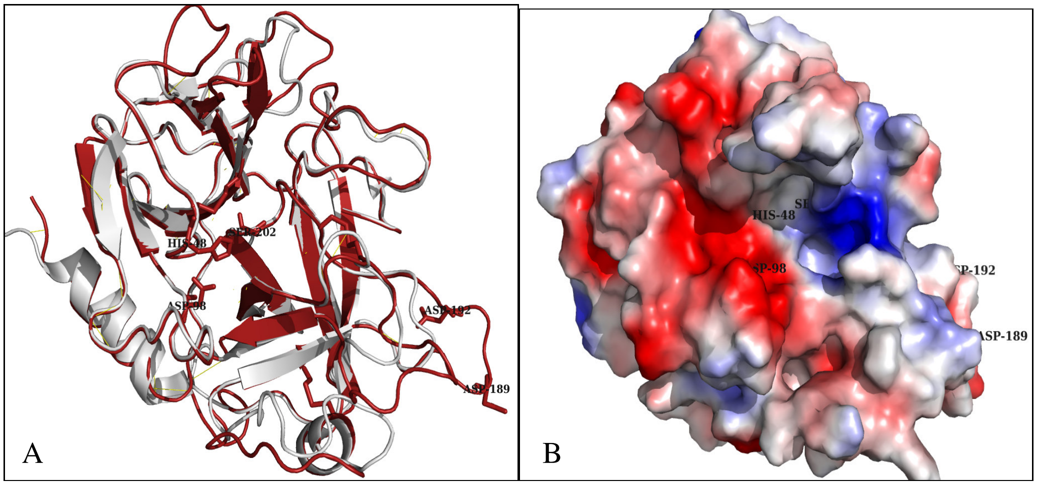


Figure 4.4.4.2: (A): Superimposition of template **1OP8** (gray) and model TVS4041-1OP8 (maroon). (B): Surface representation of model ‘TVS4041-1OP8.

The PROCHECK value was calculated to be 78.7 % and 18.6 % for favoured and allowed region that do not seems better in comparision to ‘TVS4041-1fn8’. But values obtained in ganrously allowed and and disallowed regions were 1.8 % and 0.9 % that make this model ‘TVS4041-1OP8’, better than ‘TVS4041-1fn8’ which have higher percentage of residues in non-favoured regions. The PROCHECK results are represented in ‘Ramachandran Plot’ in figure 4.4.4.2 C & D.

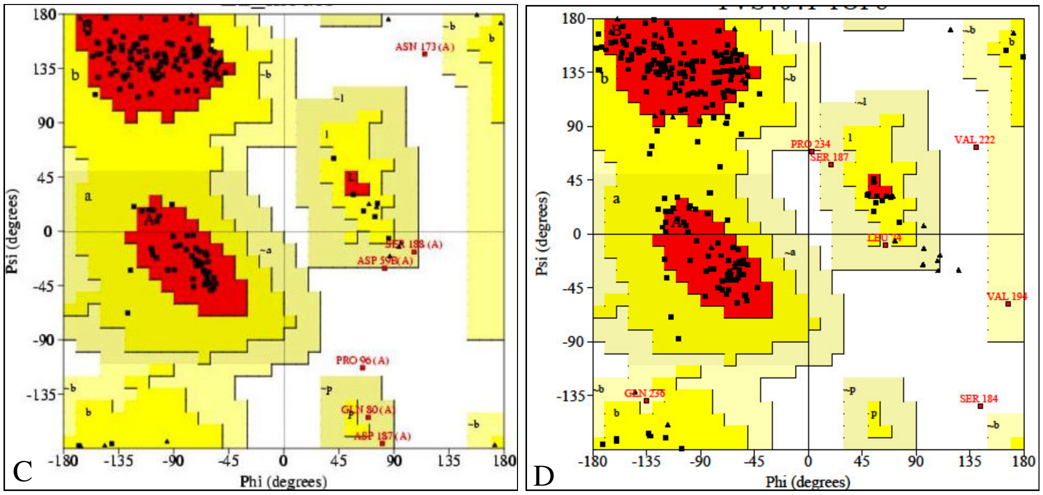


Figure 4.4.4.2: (C) Ramachandran representation of model ‘TVS4041-1fn8’. (D) Ramachandran representation of model ‘TVS4041-1OP8’. Where: red cloroured residue falls in generously allowed region or disallowed region.

4.4.5. Number of disulphide bonds:

The disulphide bonds in this particular serine type protease are only three; this is typical in lower living kingdom serine proteases. Most of the invertebrate trypsins show decreased number of disulphide bonds in comparison with the vertebral trypsins (Muhlia-Almazan, Sanchez-Paz et al. 2008). Hence it provides evidence in the evolutionary distance from inverterbral to vertebral organisms. The three conserved disulphide bonds near the active site of chymotrypsin family seem essential to preserve the structural confirmation necessary for catalytic activity. Most inverterbral trypsins posses six disulphide bonds such as bovine trypsin (bos taurus), while trypsins from lower organisms like shrimp (*P. vannamei*), resides four disulphide bridges and crayfish (*P. leniusculus*), mosquito (Anopheles gambiae) bears three disulphide bonds (Fehlhammer and Bode 1975; Muller, Crampton et al. 1993; Klein, Le Moullac et al. 1996; Hernandez-Cortes, Cerenius et al. 1999). This difference in number of disulphide bonds have been described to be linked with the early separation of vertebral and invertebral trypsins in evolutionary events (Titani, Sasagawa et al. 1983).

4.4.6. Calcium binding domain:

Ca^{2+} binding is reported to enhance the thermal stability in serine proteases which in turn increases the resistance to proteolysis (Yang, Huang et al. 2005). When looking in to the calcium binding domain from the vertebral alignment, it is not much likely that this site could be used for divalent calcium ion binding. In known reported structures such as *anionic salmon trypsin* (**1bzx**) three Glu (E) 70, 77, 80, one Asn (N) 72 one Val (V) 75 and water molecules take part in charge sharing (Helland, Leiros et al. 1998).

Many invertebrate trypsins from Fish and insect are reported to have no Ca^{2+} ion dependency on their activity (Kishimura and Hayashi 2002; Lopes, Juliano et al. 2006). Here for the invertebral aligned structure of trypsin from Fusarium oxysporum (1FN8) is not reported to have any Ca^{2+} binding site while trypsin from Streptomyces Griseus (**1SGT**), is reported to have calcium binding domain but in far distant and dispersed site as compared to common vertebral calcium binding site residue. In

1SGT Asp138, Ala151, Glu154 and Glu208 are reported to be involved in occupying of Ca^{2+} ion.

On the basis of invertebral alignment, some loops seem to have the ability to occupy Ca^{2+} ion. This region includes interdomain loop and region just after the catalytic Asp (D) residue. Interestingly, this region is very similar to the one aligned with Anopheles trypsin-1 (Swiss-Prot P35035) and contains three Glu (E) lying near to each other, making a suitable environment to compensate their charge with a positive metal ion (figure 4.4.6).

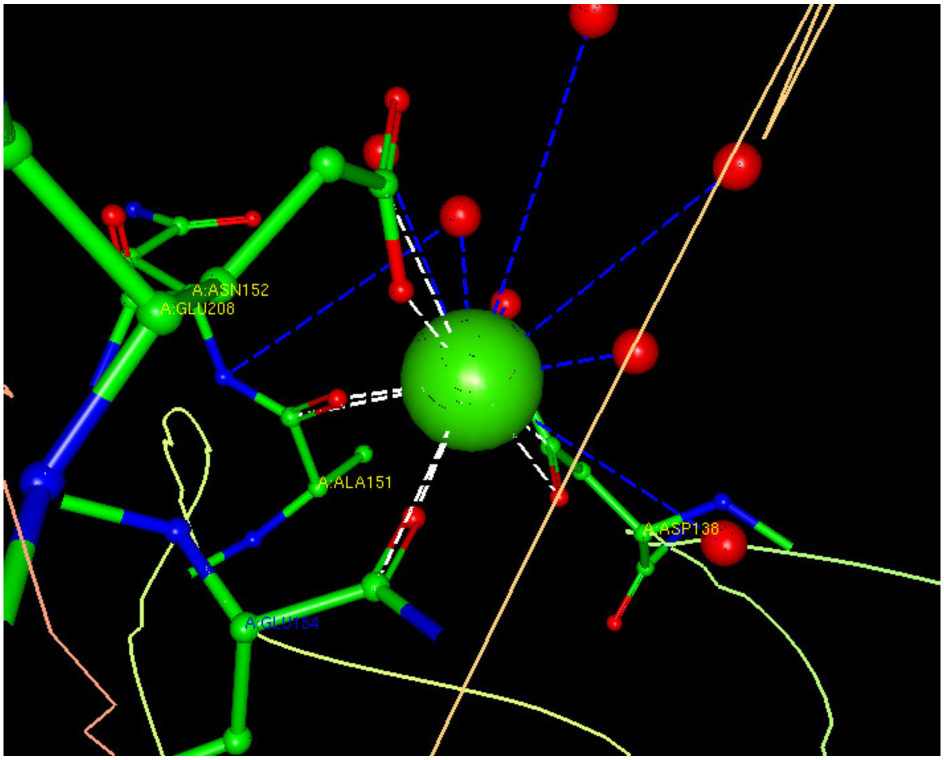


Figure 4.4.6: Calcium binding site from Streptomyces Griseus trypsin (1SGT), involve residues D138, A151, N152, E154, E208. These residues are from distant part of polypeptide, other than commonly known site, describe for higher vertebral trypsins.

4.4.7. Residues around active site:

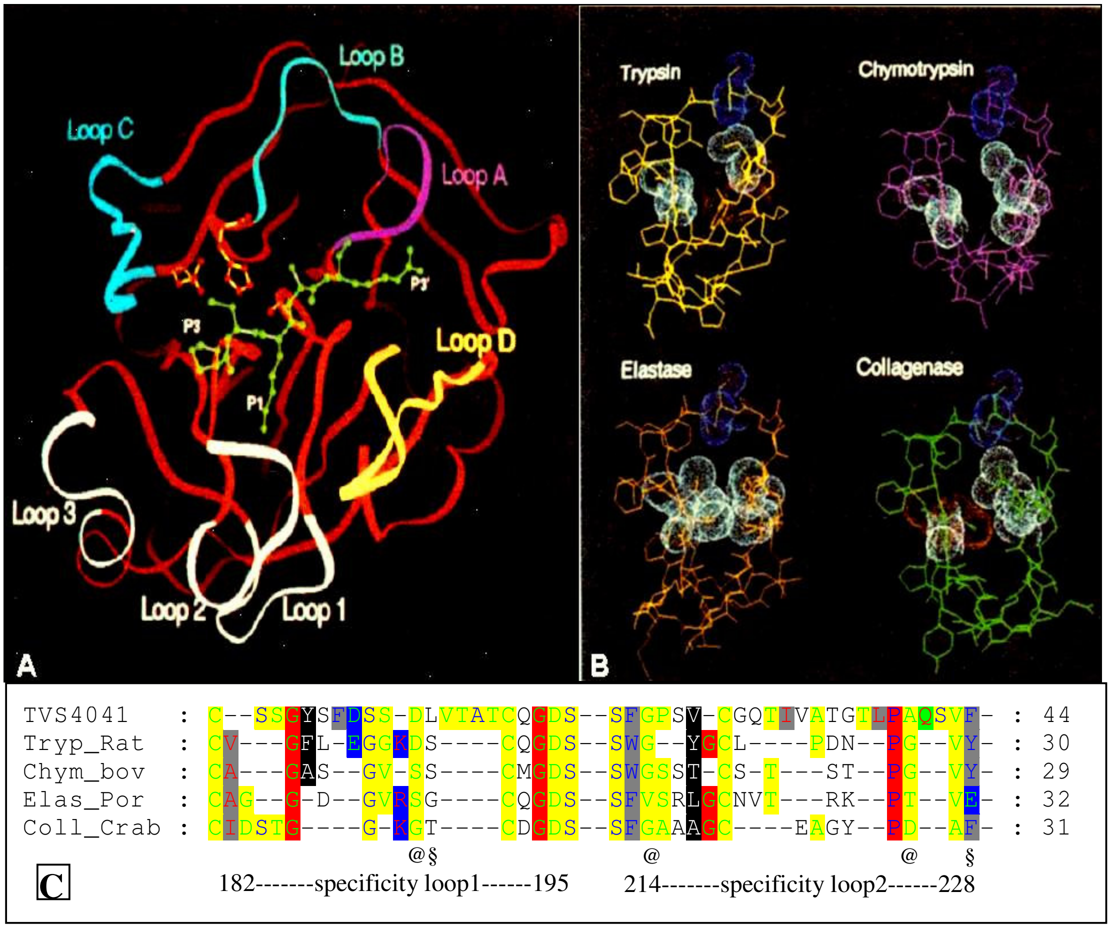
The residues around first conserved active site His (H) seem to be more in close agreement to invertebral trypsins, while the second conserved active site residue Asp (D) is more evolutionary related with vertebral type of trypsins. The right side of the third conserved active Ser (S) is more conserved to vertebral type of trypsin irrespective of the absence of Cysteine, while left side of this conserved residue is very much unique. But it is most similar to elastase 2 from *vibrio fischeri* sequence, shown in closer *spp.* alignment to site 4 (figure 4.4.2.2).

4.4.8. Substrate specificity pocket:

When looking into the sequence of TVS4041, it is difficult to predict which category does this special sequence fall. Three loops are reported to take part in substrate specificity determination for chymotrypsin-like proteases, Loop1 (182-195), Loop2 (214-228) and loop3 depicted in figure 4.4.8.1(A). These loop confirmation not only determine the capacity (volume) of P1 residue from the substrate/inhibitor, but the residue S1 from the bottom of specificity pocket determine the charge of the incoming P1 residue from substrate/ inhibitor. Based on preference for P1, the chymotrypsin-like fold is divided in four broader categories Chymotrypsin, Trypsin, Elastase and Collagenase, represented in figure 4.4.8.1(B) (Perona and Craik 1995). In order to determine the specificity of this particular chymotrypsin like domain, the two specificity determining loops were aligned from known Chymotrypsin, Trypsin, Elastase and Collagenase activity residing proteases 4.4.8.1(C).

Based on this alignment 4.4.8.1(C), it was found hard to say what will be the specificity of TVS4041, since it carries similarities with all four classes. When compared to the closer spp. alignment it is much similar to elastase of *V. fischery* with unique insertion sequence “LVTAT” in loop1 (figure 4.4.2.2, site4). But the sequences from *V. fischery,* further do not contain any Asp residue while TVS4041 sequence contains two Asp, that can maintain the negative charge inside the pocket which is necessary for trypsin like activity. When inspecting the modelled ‘TVS4041-1fn8’ [based on **1fn8** (*fusarium oxysporium*)], the presence of two Glycine on the neck of the specificity pocket further prove it closer to trypsin. But unusual to other trypsins, an extended loop is present in between the active serine and the specificity pocket (figure 4.4.8.2.A). Similar proves can be seen from the model ‘TVS4041-1OP8’, where Asp are in bottom of specificity pocket and number of Gly inside specificity pocket are making place for larger residues like Arg or Lys (figure 4.4.8.2.B).

From both the surface representation of models 'TVS4041-1fn8' and 'TVS4041-1OP8', the specificity pocket brim is covered with positive charge residues (figure: 4.4.4.1.B & 4.4.4.2.B). This character probably roleout the secondry specificity.



```
TVS4041   : C--SSGYSFDSS-DLVTATCQGDS--SFGPSV-CGQTIVATGTLPAQSVF- : 44
Tryp_Rat  : CV---GFL-EGGKDS----CQGDS--SWG--YGCL----PDN--PG--VY- : 30
Chym_bov  : CA---GAS--GV-SS----CMGDS--SWGSST-CS-T---ST--PG--VY- : 29
Elas_Por  : CAG--G-D--GVRSG----CQGDS--SFVSRLGCNVT---RK--PT--VE- : 32
Coll_Crab : CIDSTG----G-KGT----CDGDS--SFGAAAGC----EAGY--PD--AF- : 31
                         @§          @                 @  §
            182-------specificity loop1------195   214-------specificity loop2------228
```

Figure 4.4.8.1: (A): Depiction of loops that govern over substrate specificity. Loop1, 2, 3 makes the specificity pocket and determine the volume and charge condition appropriate for the substrate P1 residue. None of these residues directly contact with P1 (green), active D102, H57, S195 are on the top of this specificity pocket (yellow). Loop A, B, C, D control the secondary specificity, Loop C interacts with substrate N-terminal, loop A, D interact with the C-terminal leaving group of scissile bond. (B): Architecture of catalytic machinery of serine proteases, active ser195 is shown blue on the top, negatively charge residues (red), natural residues (white). Shape and electrostatic charge of specificity pocket determine by key residues indicted with van der waals surfaces. Hence P1 rule out for trypsin is positively charge Lys/Arg, for chymotrypsin is Phe/Tyr/Trp, for elastase is Ala and for Collagenase is any of the previously describe. (C): alignment of specificity pocket loops including TVS4041 for all four types of serine proteases, specificity residue indicated as @, secondary determinants indicated as §. [Source: (Perona and Craik 1995)]

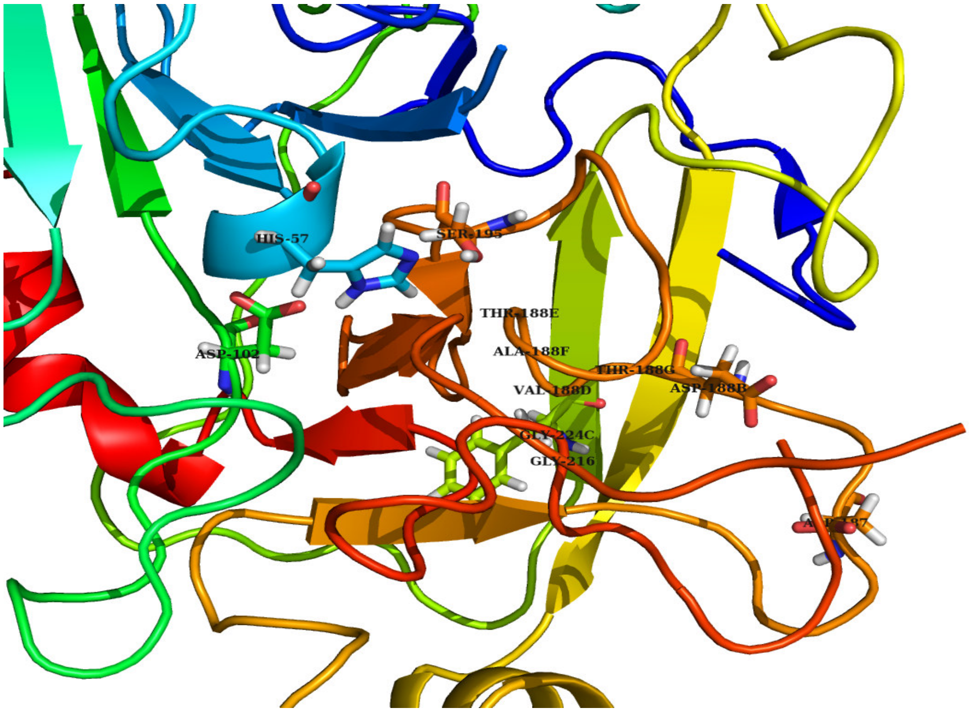


Figure 4.4.8.2 A: Modelled ‘TVS4041-1fn8’ showes the Asp inside and near the bottom of pocket, representing the Trypsin like activity, with 2 Gly around the neck of specificity pocket. The extended loop between active site residue and specificity pocket, contains VTAT on the brim of specificity pocket.

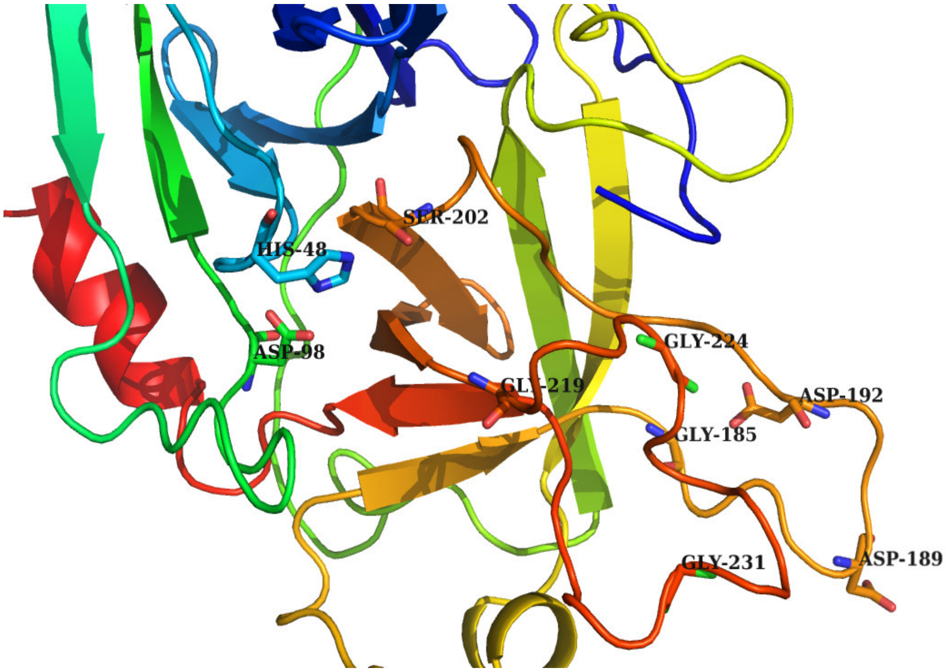


Figure 4.4.8.2 B: Modelled ‘TVS4041-1OP8’ shows the two Asp residues in the bottom of specificity pocket. Presence of number of Gly inside the specificity pocket makes the space for larger residues like Arg or Lys.

4.4.9. *C*-terminal domain:

The C-terminal domain is more protruding compared to other well known and structurally resolved serine proteases. Most of the aligned sequences from closer species contain similar unique C-terminal structures. Irrespective of the length of polypeptide the tail part of the protein is conserved with rich segment of G & S followed by R & K rich residues in extreme tail (figure 4.4.2.2 B, site7 & 8). Tripathi et al. have reported several kinds of serine protease family linked domains found in the microbial sources, used in several protein protein interactions. Examples include SI_PDZ, LON-AAA_S16, Clp, DDpept, etc. They are reported to have special reorganization function for signalling, target binding, pathogenesis, ligand binding and localization (Tripathi and Sowdhamini 2008). It could be a possiblity that this domain involves in the function of host pathogen relation.

The function of this domain as a protease inhibitor is also not out of arising queries. Since the *C*-terminal contains both Arg and Lys at the end of a long *C*-terminal tail it could have an inhibitory role, as adefensive mechanism for indigenous proteins. Prediction of TVS4041 through SecretomeP 2.0 Server (at www.cbs.dtu.dk, significance >0.5), yield 0.9484 that indicate the presence of unusual amino acid or structure that are responsible for single sequence independent non-classical protein secretion (Bendtsen, Kiemer et al. 2005). In concerning with the suspection of this domain as a secretional sequence, only the *C*-terminal part was blast into NCBInr database, that resulted in many entries related to secretional system proteins or transmembrane enchoring proteins.

These entries includede preprotein translocase subunit SecY [*Methanosaeta thermophila*], GTP-binding protein HFLX [Toxoplasma gondii VEG], putative phosphate transport system permease [Campylobacter jejuni], putative transporter subunit: membrane component of ABC superfamily [Escherichia coli UMN026, type II and III secretion system protein [Polynucleobacter necessarius], integral membrane protein [Thermus thermophilus HB8], surface antigen gene [Methanosarcina acetivorans C2A] and Abortive infection protein [Planctomyces maris DSM 8797]. The entries obtained by blast give a clear indication of this *C*-terminal part as a membrane enchore sequence that is either used to transport this protein out of cell or to keep it attached to the surface of bacteria (*v. salmonicida)*, so that bactria can invade the host by proteolytic degradation.

4.4.10. Autolysis loop:

The chymotrypsin-like proteases are prone to autolysis, in dissolved condition their activity decreases due to autolytic degradation with increasing time. This autolysis is due to the flexible surface loops, mainly onto autolysis loop (positions 141–152 in chymotrypsin numbering) and some time to the loop near the active site His57, Nβ3- Nβ4 loop (positions 59–63 in chymotrypsin numbering), the interdomain loop (positions 110–132 chymotrypsin numbering) (Mary, Achyuthan et al. 1988; Villalonga, Reyes et al. 2004).

Based on the initial characterization results described in section 4.8.0, TVS4041 is a true trypsin therefore we would look into the autolysis loop considering this target protein has affinity for Arg/Lys. TVS4041 possesses Arg 143, Lys66 and Arg 129 in these positions respectively (Figure 4.4.2.1, residues in red color). But on looking into the results obtained during purification attempts (Figure 4.10.2.C three different sizes were detected in westrn blot ~32, ~24 and ~15 KD. This represents full lenth polypeptide chain and suspected lysis on Arg 143 and lysis on Arg 129 only.

The *Streptomyces griseus* trypsin (1SGT) is known to stable against autolysis. Lee et al, have investigated this ability and found the absence of disulphide bridge around the autolysis loop. (Lee, Park et al. 2004). This is conserved only between higher vertebral sources, neither TVS4041 posses this disulphide bride. But this disulphide is also described to retain the stability of trypsin even after autolysis that has been observed in case of bovine β-trypsin. This enzyme remains active and stablizes in the form of two clipped polypeptides. They retaning shape and structure of active enzyme, due the presence of that conserve disulphide bond around autolysis loop (Walsh 1970). Hence, in case of autolysis in TVS4041 this protein is predicted to be unstable, since it do not have disulphide bond around autolysis loop.

Lee et al, have further pointed out the presence of a strong bend in the autolysis loop just beside the valunerable Lys145. This bend occurs due to salt bridge between E146 and R222, leading autolysis prone Lys out of the reach of catalytic triad (Lee, Park et al. 2004).

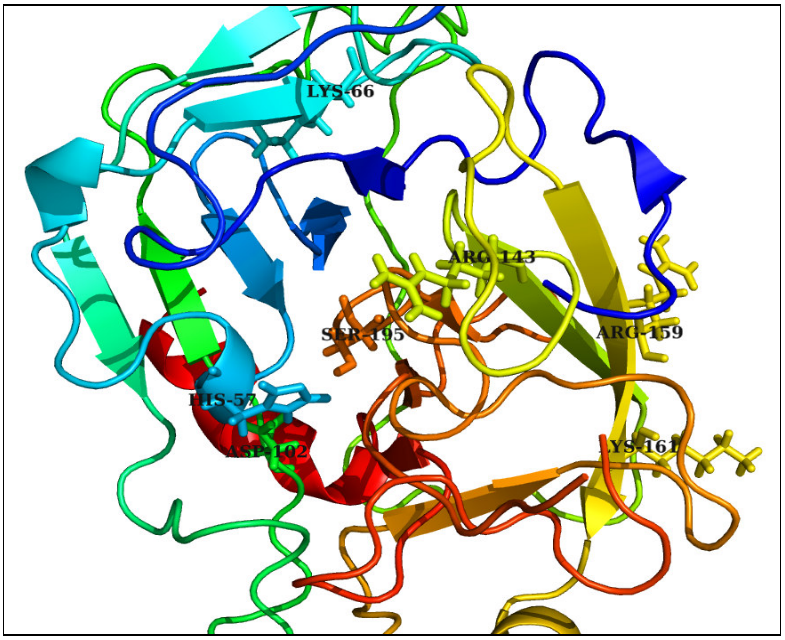


Figure 4.4.10: Presence of Arg/Lys in the autolysis prone loops in model 'TVS4041-1fn8'. Arg 143 is bend invert near the active serine 195. Arg 159 and Lys 161 from Nβ5 loop are exposed to surface and prone to intra digestion.

4.4.11. Residue for cleavage of active protease:

In the alignment with closely related species, it can be seen that the residue prior to the mature TVS4041 is Tyr, which is common in most of vibrio species (Figure1.4.2.2.A; site1). Hence, here it is likely that the activation of this enzyme is by chymotrypsin cleavage of Tyr (Y) prior to 'IVGG' (conserved *N*-terminal sequence).

4.4.12. Loops:

The loops in all invertbral serine proteases are very irregular in length. Hence, a similar pattern is only conserved around catalytic residues or in evolutionary related species. The longer loops and the lower number of Pro, increased Gly contents are reported for cold adapted enzymes (Siddiqui and Cavicchioli 2006). Alignment 4.4.2.1 is representative of this phenomenon in case of TVS40401.

4.4.13. Surface Hydrophilicity:

Surface hydrophilicity, particularly negative charges is the unique feature found for almost all of the reported cold adapted enzymes like trypsins, beta lactamse, subtilisins, malate dehydrogenase (Siddiqui and Cavicchioli 2006). At lower temperature, water has higher viscosity and higher surface tension that can disrupt the hydrogen bonds (Kumar and Nussinov 2004). In cold adapted enzymes, the higher dielectricity of water is compensated by greater surface charge accumulation. Charged and polar amino acid helps in better salvation and flexibility of these enzymes in unfavourable cold tempature (Kumar and Nussinov 2004).

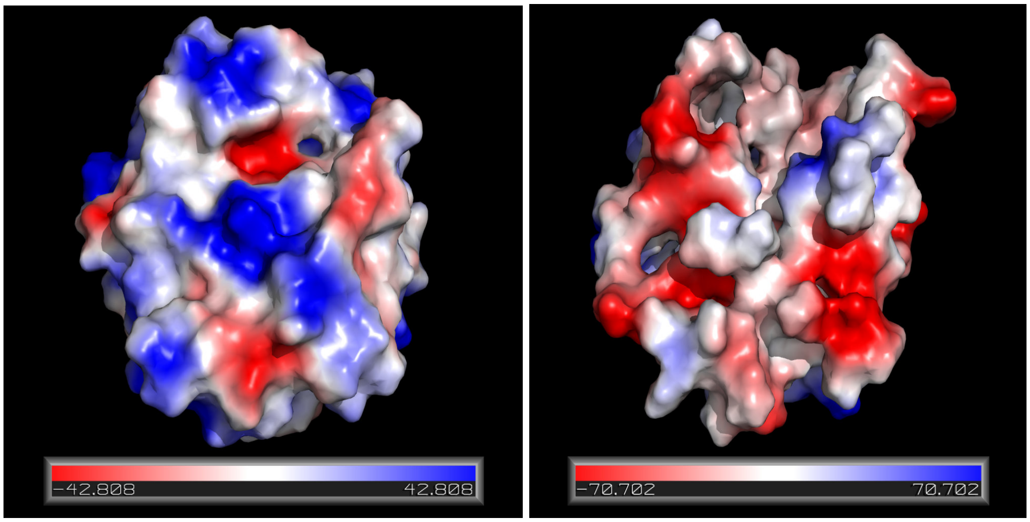


Figure 4.4.13: surface charge of template 1fn8 (left hand side), in comparision to modelled 'TVS4041-1fn8' (right hand side).

4.4.14. Methionine as aquatic compatibility feature:

Numbers of methionines are believed to be higher in marine organisms. Among all type of trypsins M104, M180 is most conserved while M135, M145, M175, M242 is additional in Coldfish trypsin (1AOJ) (Leiros, Willassen et al. 1999). TVS4041 has 5 Met, in alignment (figure 4.4.2.1); where it can be seen that none of them are well aligned. Hence it can be said in spite of being from an aquatic source there are some essential differences between higher and lower organisms. In figure 4.8.0, TVS40401 is shown to have a greater surface potential change than its template

1FN8, that is from mesophilic source. This is similarly high as that found for shrimp alkaline phosphatase (-80) (de Backer, McSweeney et al. 2002).

4.4.15. Aliphatic residues:

Aliphatic resides (Ala, Leu, Ile, Val) are known to improve the thermostablity of proteins (Ikai 1980). When compared the amino acid composition of aligned segment with some well-studied trypsins (pdb code 1HJ8, 1AOJ, 1DPO, 1BTP), the increased number of aliphatic residues can be detected (figure 4.4.2.1). For reference, the aliphatic residues of TVS4041 was compared with other trypsins from the previously conducted work from our research group (Leiros, Willassen et al. 1999).

Table 4.4.15: Percent variation of aliphatic residues from vertebral trypsins and TVS4041; source (Leiros, Willassen et al. 1999)

	TVS4041 (%)	Higher vertebrate (%)	Cold fishes (%)	Other fishes (%)
Ala	**9.1↑**	6.67	6.4	6.7
Ile	**5.7↓**	6.8	4.5	6.62
Leu	**7.52↑**	6.2	6.21	5.95
Val	**9.1↑**	7.83	8.78	7.74
total	**31.42↑**	27.5	25.89	27.01

From the above compared values, it is obvious that TVS4041 has increased aliphatic residues then cold fishes, even in comparison with mesophilic and vertebral trypsins. Hence, it can be concluded that this protein can be well tolerated in case of increasing temperature. Section 4.7.7 is an evedience of this statement. Since, TVS4041 was attempted to expresse in the range of 15, 22, 30, 37 and 42 °C, but the expression and stability over time remains highest at 37 °C, in comparision to lower tempratuers.

4.4.16. Overall charged residues:

In connection with charged residues in TVS4041, the pattern was compared with the pre-evaluated large set of database in the aligned region (Leiros, Willassen et al. 1999). Asp residues found highest in number for TVS4041 and same was seen for the cold adapted fishes. But no significant change could be observed in Glu, from the other comparative data. While on the other hand all of the positively charged residues are much lower innumber in comparision to other vertebrate, fishes and cold fishes.Hence overall negative charge is much higher in TVS4041 that is also the case for cold adapted fishes.

Table 4.4.15: percent variation of charged residues from vertebral trypsins and TVS4041; source (Leiros, Willassen et al. 1999)

	TVS4041 (%)	Higher vertebrate (%)	Cold fishes (%)	Other fishes (%)
Asp	**5.3↑**	3.85	4.28	3.94
Glu	3.4	3.44	3.92	3.13
Total-ve	8.7	7.29	8.2	6.07
His	**1.5↓**	1.97	3.06	2.46
Arg	**1.9↓**	1.97	2.79	2.82
Lys	**2.3↓**	4.2	3.37	2.37
Total+ve	**5.7↓**	8.14	9.22	7.65
Net charge (R+K-D-E)	**-4.5↑**	-1.12	-2.04	-0.88
Arg/Lys	0.82	**0.47**	0.83	1.19
Arg/(Lys+Arg)	0.45	**0.38**	0.44	0.55

Some of the cold adaptation criteria for lower Arg/Lys ratio, Arg/(Arg+Lys) ratio does not match very well in this particular data set (Leiros, Willassen et al. 1999; Adekoya, Helland et al. 2006).

4.4.17. Higher Threonine content:

In comparison with data from previously describe sources, a significantly higher quantity of Thr was observed in the aligned region (Figure: 2.2.4.1). Thr is a polar residue that tends to appear on the surface of protein. In TVS4041, it mostly appears in loop regions. The probable role of threonine seems to increase the surface charge.

Table 4.4.15: Percent variation of aliphatic residues from vertebral trypsins and TVS4041; source (Leiros, Willassen et al. 1999)

	TVS4041 (%)	Higher vertebrate (%)	Cold fishes (%)	Other fishes (%)
Thr	**8.9↑**	3.98	4.27	4.48

Threonine is a naturally occurring β-branched residue that have less freedom of confirmation, therefore it impose reduce flexibility of the structure, hence increases the rigidity of structure. Increased Thr contents could be the nature's compensation for reduced rigidity due to the lack of disulphide bonds in comparision to vertebral serine proteases.

4.5.0. Cloning of the targets:

In progress of cloning through gateway, for all selected targets almost all had been cloned into the entry clone, provided they have been confirmed through sequencing for correct frame and absence of mutation.

Table: 4.5.0: Status of targets cloning, upto the entry clone formation.

Entry clone in progress	LexA			RadA			ThiJ			HslV			Trypsin			Collagnase			ProtinaseK			ToxR		
	I	II	III	I	II	III	I	II	III	I	II	III	I	II	III	I	II	III	I	II	III	I	II	III
Gene replication (*PfX* polymerase, invitrogen®)	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Replication with att site (attB1 & attB2 primer)	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Gene insertion in attP vector (pDONOR 221)KmR	√	√	√	√	√	√	-	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Transformation in *E. coli* DH5α cells (Entry clone)	√	√	√	√	√	√	-	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	X	√
Kanamycin resistant colony isolation	√	√	√	√	√	√	-	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	-	√
Plasmid mini isolation (QIAGEN™ Kit)	√	√	√	√	X	√	-	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	-	√
Gene conformation (with gene specific primer)	√	√	√	X	-	X	-	-	√	√	√	√	√	√	√	√	√	√	√	√	√	-	-	-

I =Native Protein; II = N-terminal His Tag; III = C-terminal His Tag,

4.6.0 Test Expression studies:

For test expressional studies, three targets have were successfully expressed and confirmed through Tandem Mass Spectroscopic analysis. Their progresses have been summarized in the tables 4.6.0.A, B, C.

LexA and HslV gave high yield and their solubility was good. While, in case of TVS4041 several trials and attempts for expressions did not results in detectable expression. Since in these expression trials, signal sequence was selected from the *A. salmonicida,* therefore it could not be recognized by E. coli indigious system. Hence expressed protein did not passed to periplasm and could not be folded and hence subjected to cytoplasmic degradation.

Table: 4.6.0.A: status of expression clone formation and pilot expression of LexA. Where AI stands for BL21 (DE3) AI, CP stands for BL21 (DE3) CodonPlus chemically competent expression cells. *ND=not determined, X=no detection*

Expression clone in progress	**LexA**											
	Native Construct		***N*-terminal Tags**								***C*-terminal Tags**	
Entry clone plasmid x Destination Vectors (different fusion tag attachment)	pDEST 14 (no tag)		MBP	Gb1	Gst	NusA	Z	Gst	Trx	6x His	pET DEST 42	
Tran. in *E. coli* cells (BL21: AI, DE3, Codon Plus)	AI	CP	CP	CP	CP	CP	CP	CP	CP	CP	AI	CP
Ampcillin resistant colony isolation	√	√	√	√	√	√	√	√	√	√	√	√
Plasmid mini isolation (QIAGEN Kit)	√	√	√	√	√	√	√	√	√	√	√	√
Gene confirmation (with gene specific primers)	√	√	√	√	√	√	√	√	√	√	√	√
Sequencing with vector primers	√	√	√	√	√	√	√	√	√	√	√	√
Small Scale Protein expression test	√	√	√	√	√	√	√	√	√	√	√	√
Expression confirm with MS analysis	ND	ND	√	√	X	√	√	X	√	√	ND	ND

Table: 4.6.0.B: Status of expression clone formation and pilot expression of HslV. Where AI stands for BL21 (DE3) AI, CP stands for BL21 (DE3) CodonPlus chemically competent expression cells. *ND=not determined.*

Expression clone in progress	**HslV**					
	Native Construct		**N-terminal Tags**		**C-terminal Tags**	
Entry clone plasmid x Destination Vectors	pDEST 14		6x His		pET DEST 42	
Tran. in *E. coli* cells (BL21: AI, DE3, Codon Plus)	AI	CP	AI	CP	AI	CP
Ampcillin resistant colony isolation	√	√	√	√	√	√
Plasmid mini isolation (QIAGEN Kit)	√	√	√	√	√	√
Gene confirmation (with gene specific primers)	√	√	√	√	√	√
Sequencing with vector primers	√	√	√	√	√	√
Small Scale Protein expression test	√	√	√	√	ND	ND
Expression confirm with MS analysis	√	√	ND	√	ND	ND

Table: 4.6.0.C: status of expression clone formation and pilot expression of TVS4041. Where AI stands for BL21 (DE3) AI, CP stands for BL21 (DE3) CodonPlus chemically competent expression cells. *X=no detection.*

Expression clone in progress	**TVS4041**					
	Native Construct		**N-terminal Tags**		**C-terminal Tags**	
Entry clone plasmid x Destination Vectors	pDEST 14		6x His		pET DEST 42	
Tran. in *E. coli* cells (BL21: AI, DE3, Codon Plus)	AI	CP	AI	CP	AI	CP
Ampcillin resistant colony isolation	√	√	√	√	√	√
Plasmid mini isolation (QIAGEN Kit)	√	√	√	√	√	√
Gene confirmation (with gene specific primers)	√	√	√	√	√	√
Sequencing with vector primers	√	√	√	√	√	√
Small Scale Protein expression test	√	√	√	√	√	√
Expression confirm with MS analysis	X	X	X	X	X	X

4.7.0. TVS4041 Cultural condition optimizations

Initial attempts were made to express the TVS4041 through the gateway cloning system, taking the native signal sequence of cloned gene from *Aliivibrio salmonicida*. But these attempts in different vectors and various expression strains resulted in no expression at all. Then an expression system were tried to express TVS4041 gene into the pBAD/gIII vector that utilize GIII (18 amino acid sequence from flamentous phage ***fd***) as signal sequence and express in TOP10 (*E. coli*) cells. These attempts resulted positively with expressional detection and confirmation through MS analysis. Hence, successful pBAD/gIII (TOP10) system was taken to optimize further for growth conditions and to evaluation of the constructs performance.

4.7.1. Optimization of the TVS4041 expression

Initial Experiments was done with two native constructs with prosequence (N1) and without prosequence (N2). The following results show that N1 is good in yield in comparison with N2 construct. In selected two temperature conditions, the higher temperature (37°C) seemed to be better for target protein expression in comparison with lower temperature (25°C) yield (figure: 4.7.1). The results were confirmed with Mass Spectroscopic (MS) analysis.

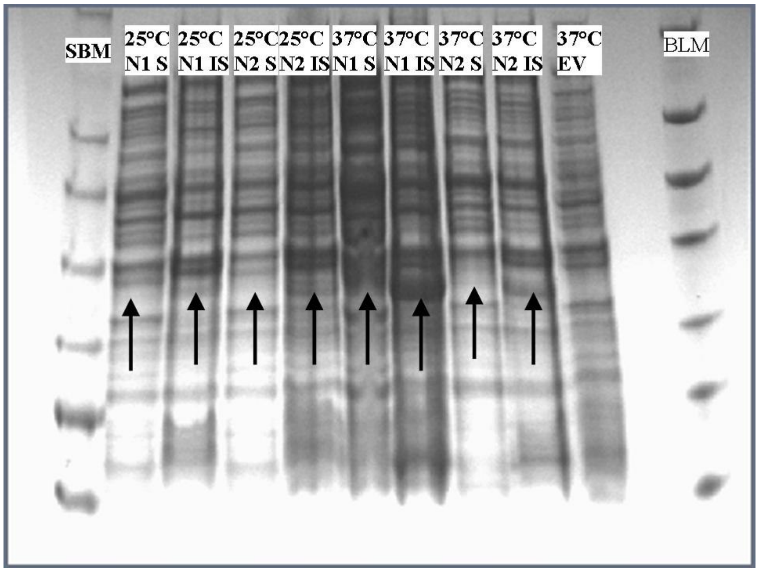


Figure 4.7.1: expression with two native constructs, with pro-sequence(N1), without prosequence(N2). Where SBM=SeeBluePlus marker, S=soluble fraction, IS=insoluble fractions from respective constructs grown at 25°C & 37°C.

For the optimization of different influencing parameter the N1 His tag construct (H1) was mostly used for the easy detection and evaluation of experimental parameters. Since the protein under investigation is suspected to be trypsin with higher hydrolysing affinity for Lys and Arg, and could be opportunistic affinity to other amino acids (Grzesiak, Helland et al. 2000). So, it is possible that this protease can autolyze its C-terminal part, which is enriched with two Arg and one Lys and in turn cleave off the poly His tail (**RK**S**R**AS*VDHHHHH**) provided by the vector as a purification and detection tag. Therefore, in all of the following experiments SDS-PAGE was also run beside the western blot to avoid the false negative results.

4.7.2. Optimal cell growth stage for induction:

Optimization experiments start with optimization of induction stage for TOP10 cells. Three major points are reported in the literature for induction of cultures, initial log phase, late log phase and initial stationary phase (Novagen 2003). Here experimental evidences in case of TOP10/pBADgIII system suggest the OD_{600} should be 1.0 for greater performance that corresponds to late log phase (figure: 4.7.2).

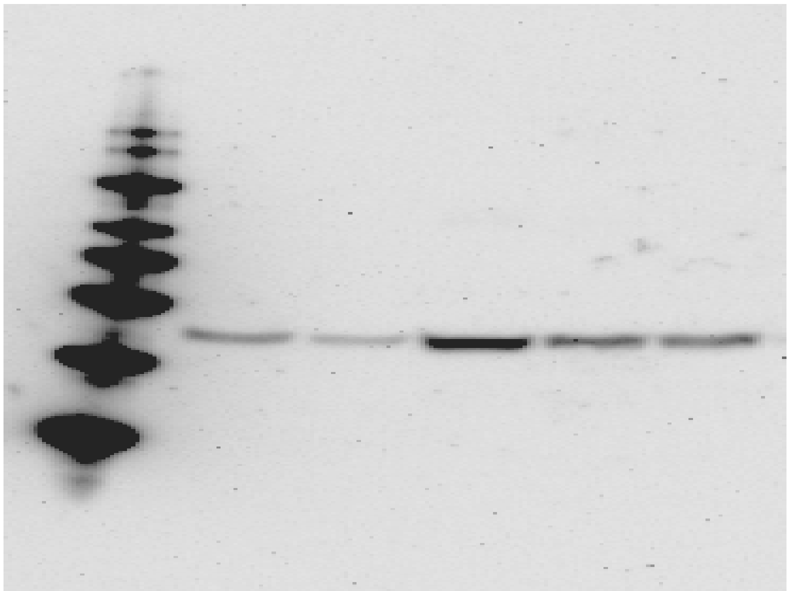


Figure 4.7.2: Western blot detection of target protein induced at different cell density levels. (From left to right XM=XP Magic Marker, OD_{600} for cell culture induction is labeled as 0.6, 0.8, 1.0, 1.2 & 1.3)

4.7.3. Variation of Inducer Concentration for increased soluble yield:

Optimization of the inducer concentration is essential in expression studies. Tuner et al. has reported that optimal concentration of inducer need to be optimize for maximal output from each batch of expression (Turner, Holst et al. 2005). On the other hand, this optimization is crucial as in some cases of over expression with strong promoter, higher concentration of inducer can lead to accumulation of inclusion bodies and aggregation (Baneyx 1999).

In the search for finding the optimal concentration of inducer (L-Arabinose) for TOPO10/ pBADgIII (invitrogen®) expression system, no considerable yield difference could be seen above the 0.2% final concentration. At 0 % L- arabinose, leakey expressions occurs which is hardly visible on western blot or SDS PAGE, while at final concentration of 0.1% expression has increased but it was less than in comparison with the final concentration of 0.2% L-arabinose. Further on, with increasing L-Arabinose no considerable soluble expression was seen (figure: 4.7.3).

Since all the experimental treatments were subdivided from the same cultural flask, therefore all of the treatments were in same homogenous stage of growth and cell density. Therefore it can be concluded that the optimal concentration of inducer (L-arabinose) for TOPO10/ pBADgIII (invitrogen®) expression system, is a function of the cell density that needs to be expressed. Furthermore, higher concentration of inducer did not have any role in aggregation or insolubility of this particular protein in the specified system.

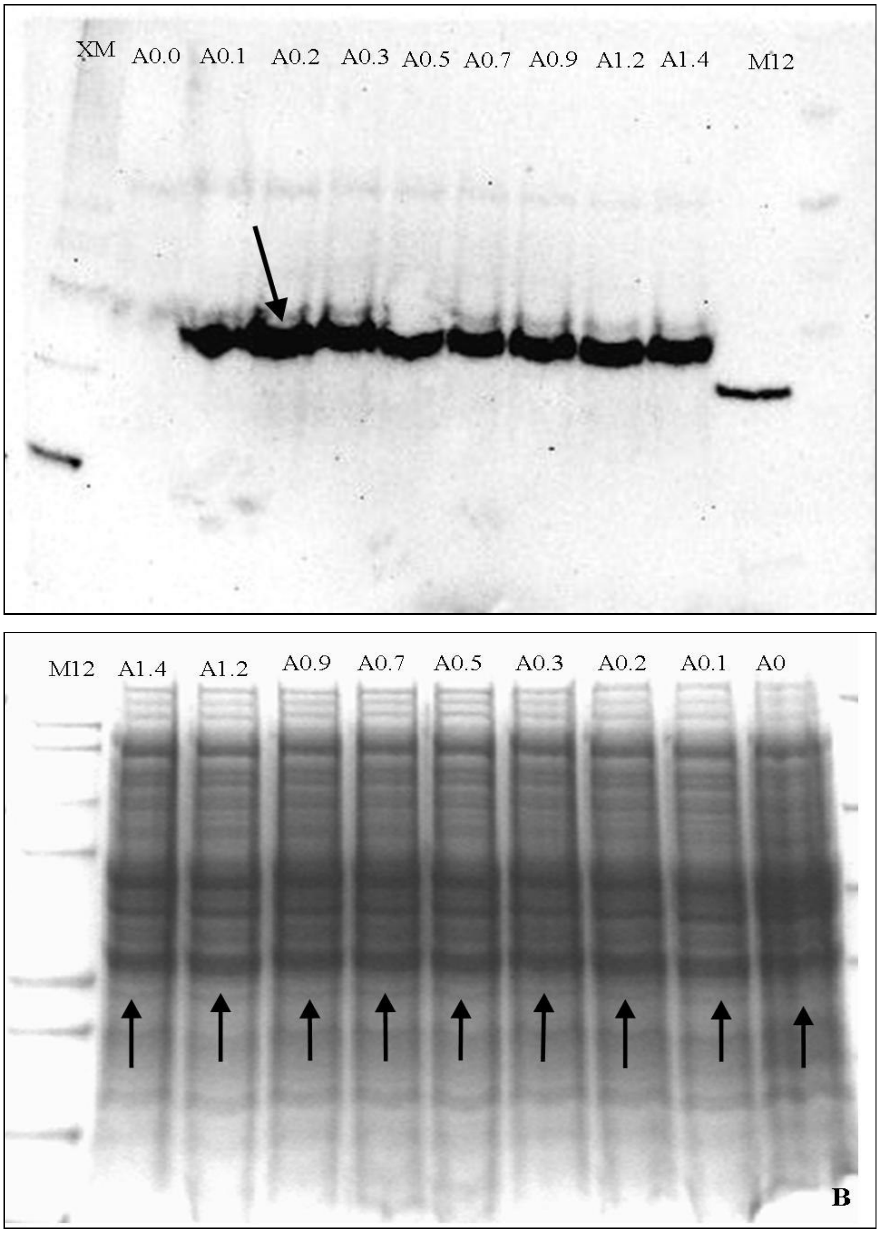


Figure 4.7.3: effect of varying Arabinose concentration in %(w/v) for increased soluble expression XP stands for Magic marker XP, M12 stands for Mark 12 protein standard. (A: Western Blot; B: SDS-PAGE)

4.7.4. Time scale parameter evaluation for soluble expression:

An experiment was run to see the solubility of TVS4041 expression with increasing time intervals. The experimentation showed that this particular protein expression seems to start after two hours of induction with maximum soluble amount at 3hrs. Harvesting of the culture after an over night induction (16 hrs) resulted in diminishing soluble protein (figure: 4.7.4).

There could be many possible reasons for the decrease in expression of target protein in an overnight induction.

i) The decrease of soluble expression could be due to the growing aggregation that might be enhanced with the increasing time period; this factor has been studied and describe in section 4.7.8.

ii) Another possibility is the escape of soluble product into the medium, which has been studied in the section and concluded in section 4.7.9.

iii) Plasmid instability is another reason that might cause the declining product yield over increasing time. Use of glucose has tested in this perspective in section 4.7.10.

iv) Plasmid loss or plasmid-less culture might be a reason for target protein loss. An attempt have been made (section 4.7.11), to check if increasing antibiotic concentration can increase the yield.

v) It is possible that expression of target protein is declining due to degradation by *E. coli* indigenous system or by autolysis through the target protease. In section 4.7.11. the result of an attempted experiment with serine type of proteases inhibitor beta lactum (ampicillin/ carbencillin) have discussed.

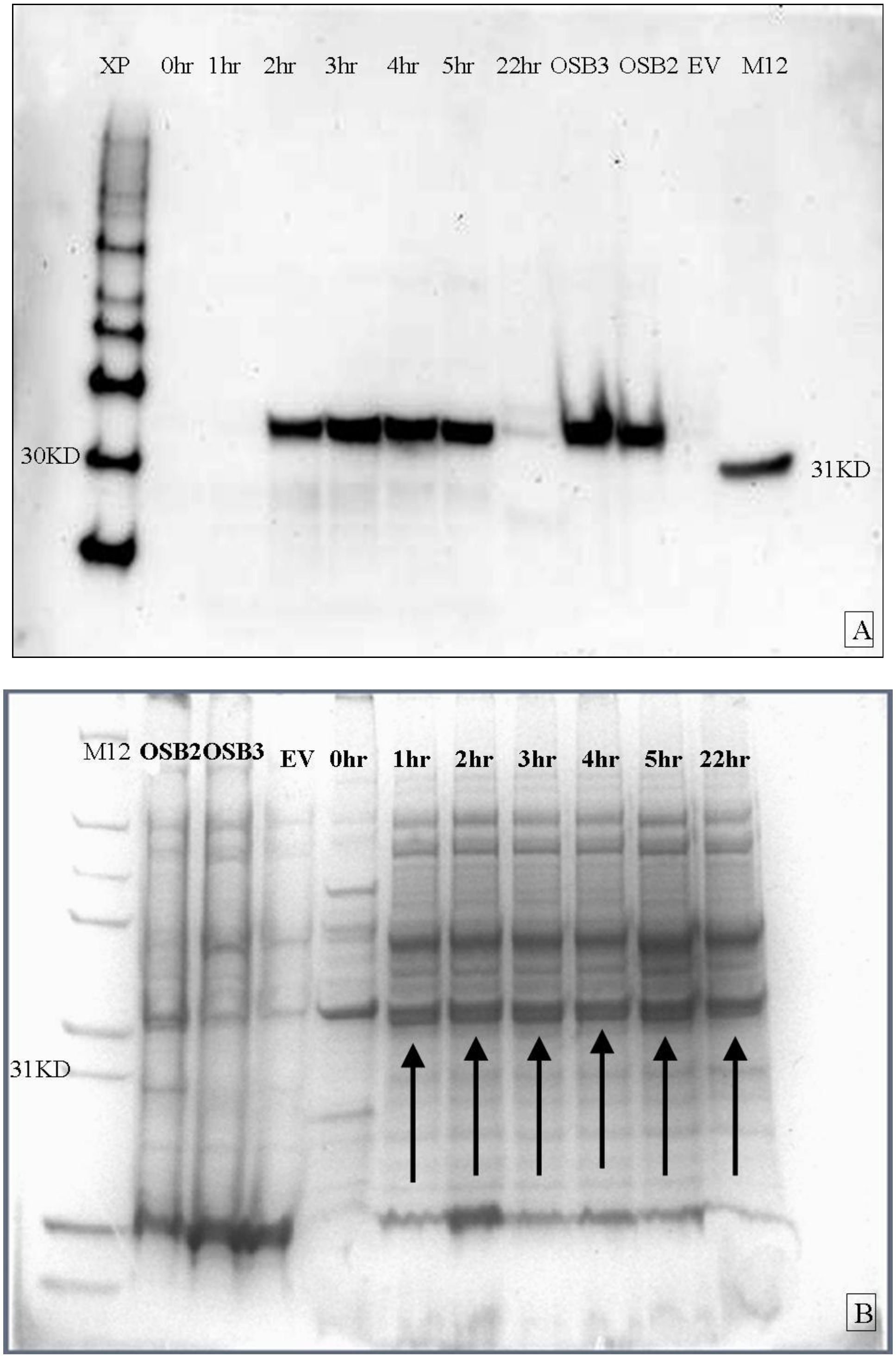
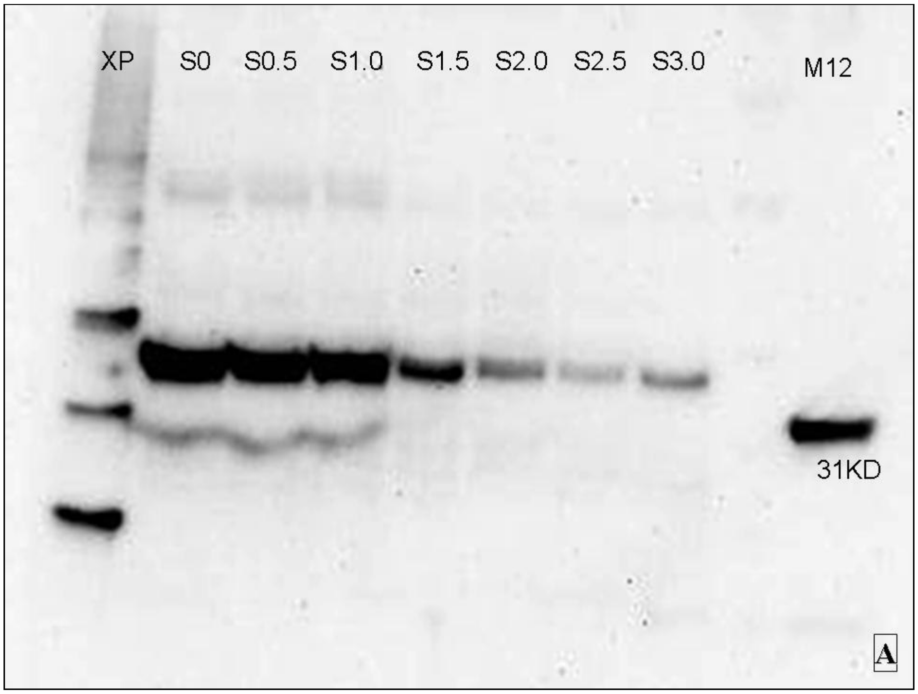


Figure 4.7.4: detection of target protein expression harvested on increasing time scale. Where, time of harvest after induction 0, 1, 2, 3, 4, 5 hrs & over night, M12= Mark12, XP=XP Magic Marker, OSB2 & OSB3 periplasmic fractions isolations after 4 hrs, EV= periplasmic fractions from empty vector. (A: Western Blot; B: SDS-PAGE)

4.7.5. Salt variation evaluation:

Based on the results from section 4.7.4, in most of the further experimental cultures were harvested after 3hrs, except where longer term effects needed to be evaluated. The protein under investigation is from a moderately halophilic bacterium *A. salmonicida* (Colquhoun, Alvheim et al. 2002). Previously isolated proteins from this organism was also found slightly halophilic up to the range of 0.6M for their optimal activity (Niiranen, Altermark et al. 2008). The protein under investigation has also evediently prove the highest activity in the presence of salt (observe during purification, section 4.10.2; figure 4.10.2.C).

The pattern obtained in the western blot analysis of the expressed cultures shows the increasing stability toward low salt concentration in cell culture. And with increasing salt their stability starts diminishing probably due to the increasing autocatalytic efficiency (figure: 4.7.5). Hence it is concluded that salt should be kept low in the modified media to avoid TVS4041 self digestion.



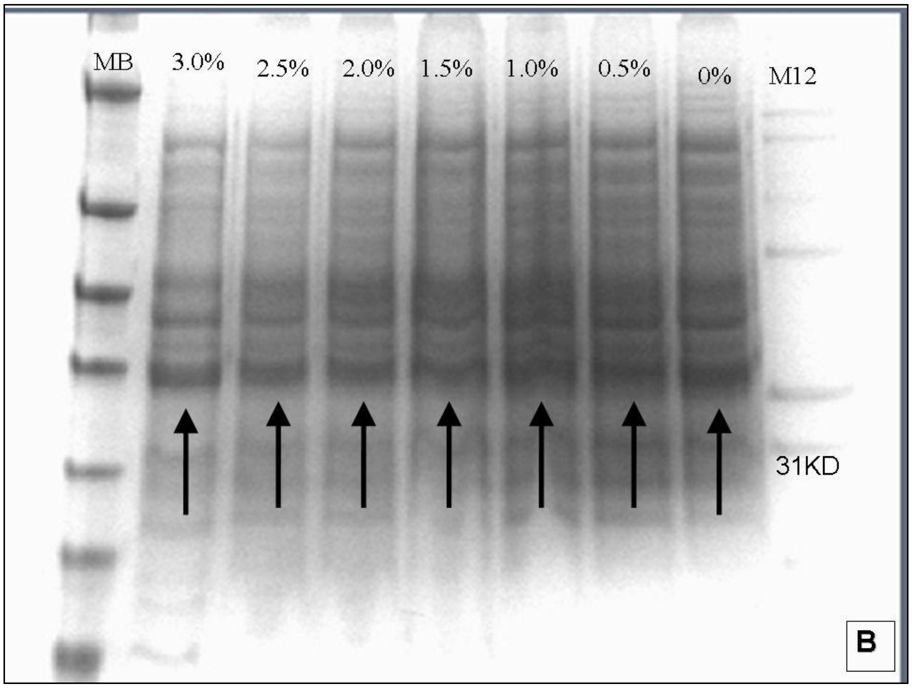


Figure 4.7.5: Effect of salt on solubility of TVS4041. Concentration of NaCl given in percentage (g/100ml). MB stands for BioLabs protein standard, M12 stands for Mark 12 protein standard, XP stands for XP Marker. (A: Western Blot; B: SDS-PAGE)

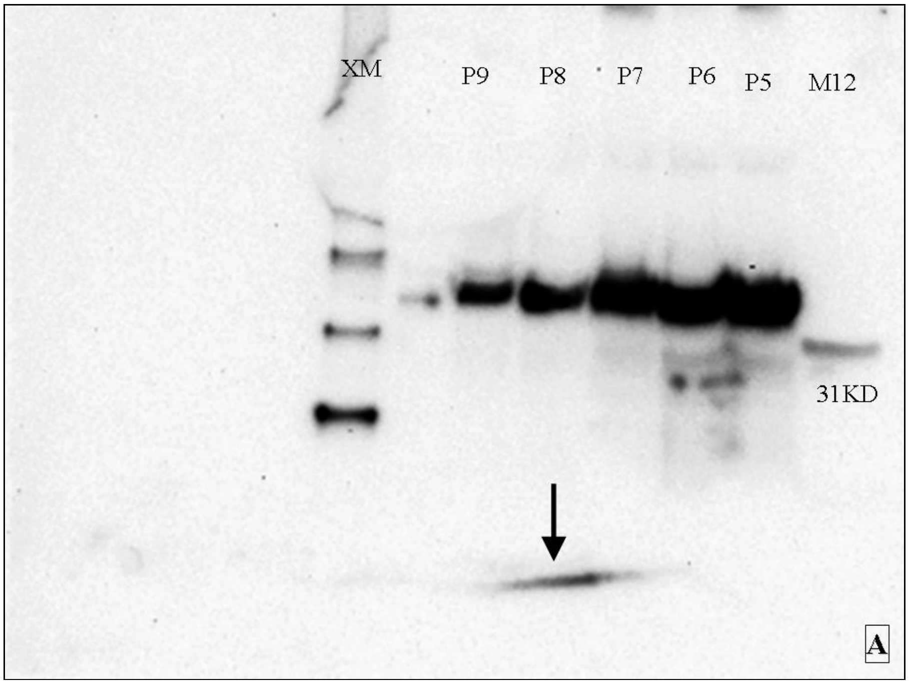
4.7.6. pH variation evaluation:

Interesting results have been observed in the experiments with the effect of varying pH of the medium. An inclining pattern was obtained in the 5 hrs expression from pH 9 towards pH 5 (figure 4.7.6). There are several reports, where expression of certain proteins up-regulated *in vivo* due to the low pH. Some of the reasons found were; influence in promoter operation mechanism, effects on transcriptional regulation and effects on signal transduction mechanism. Beside genetic regulations, gene expression profiling has also detected alterated in amino acid metabolism, transporter function, modification of cell membrane structure, and oxidative stress protection (McGowan, Necheva et al. 2003; Wilkins, Beighton et al. 2003; Leaphart, Thompson et al. 2006).

All of the above mentioned research work was done with the indigenous host of protein. Kim et al., have cloned the gene of an acidophilic protein **laccase** from Mushroom (*Coprinus congregatus)* into an *E.coli JM.* They have found that this acidophilic protein expression still depends on the acidic environment as it was required in it's indigenous host (Kim, Leem et al. 2001). Based on Kim and co-workers results it can be suspected that TVS4041 expresses under the influence of low pH in *Vibrio salmonicida*.

Being a serine type of protease, it is very much essential to find out the factors where this protein is more stable and less autocatalytic, so that greater yield can be obtained in the form of intact protein. Most of the trypsins are less active at lower pH, hence can be less autocatalytic. The results obtained in case of VST (TVS 4041) indicates similar results (Data present in characterization section 4.8.0). Thus the inability of trypsin to be active at lower pH, can be exploited during the purification procedure for stability and storage of the target protein.

Interestingly western blot detection has indicated a band in the lower molecular weight range in the bottom of pH 8 sample well. It is suspected that this is a degraded by product from the TVS4041. Since pH 8 is the optimal activity range for this enzyme as seen during TVS4041 characterization (section 4.8.0). Based on the results from section 4.7.5, and 4.7.6, media (2xYT) were further modified by adding no salt in it and by adjusting the pH6 instead of 7.4.



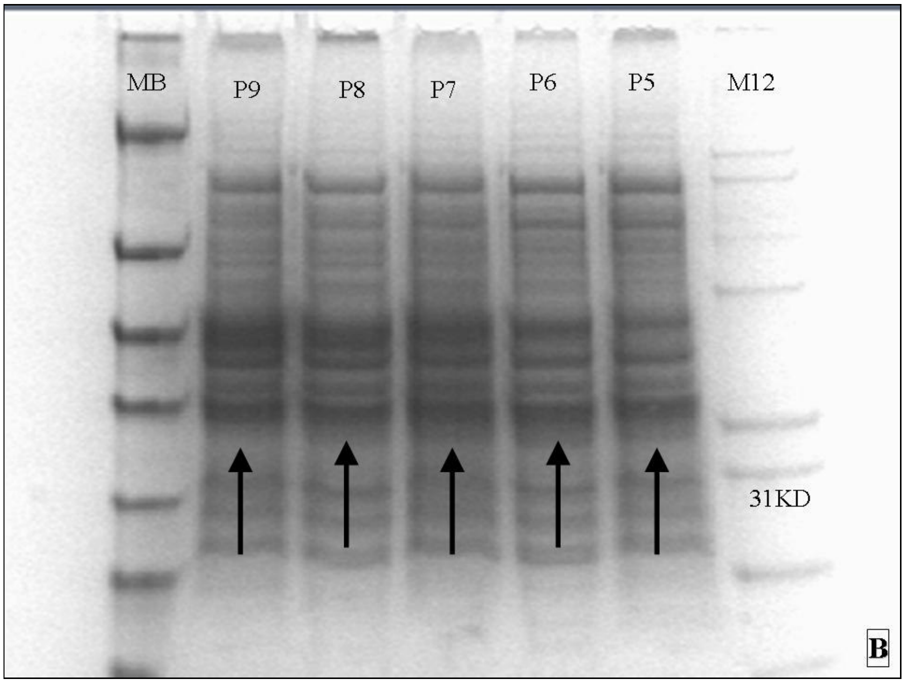


Figure 4.7.6: Detection for the effect of medium pH on soluble expression of TVS 4041. Where P = pHs (9, 8, 7, 6, & 5), MB stands for BioLabs protein standard, M12 stands for Mark 12 protein standard. (A: Western Blot; B: SDS-PAGE)

4.7.7. Effect of temperature on protein expression:

An experiment with modified media (2xYT with no salt added, pH6) was designed to see the effects of temperature on the soluble recombinant protein production, after 3hrs and after 18 hrs. Among all of the treatments, 37°C was found best for both 3hrs and 16 hrs expressions (figure: 4.7.7). For the problem of insolubility it is stated that expression at low temperature with prolonged time is useful but here this experiment shows that in case of TVS4041 prolonged incubation at lower temperature such as 30°C, 22°C and 15°C was not proved successful in greater soluble yield production. Temperature effects rate of metabolic mechanism of bacteria (TOPO10) and to the recombinant protein production as well. It had observed that at higher temperature (37°C), recombinant protein production had enhanced many folds overall.

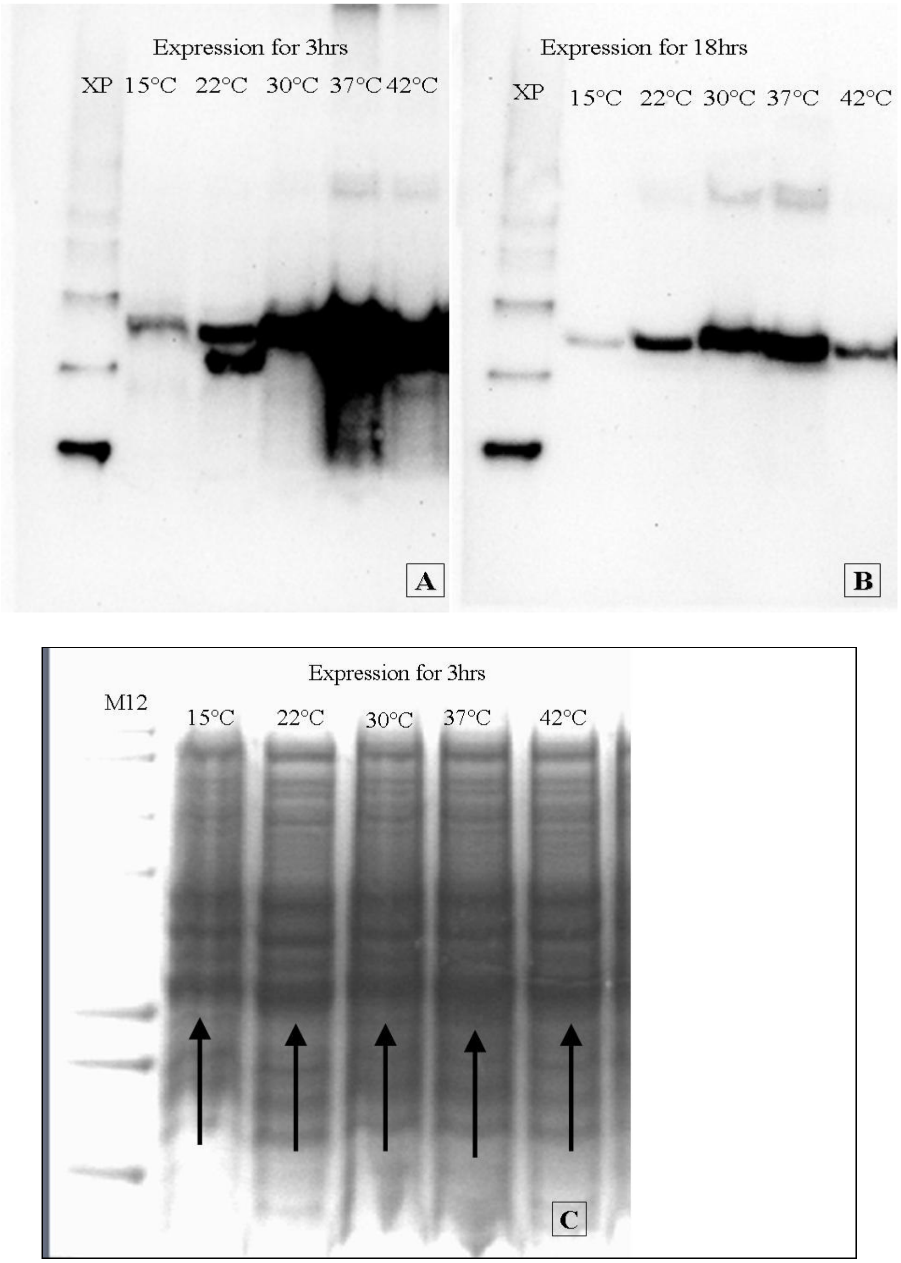
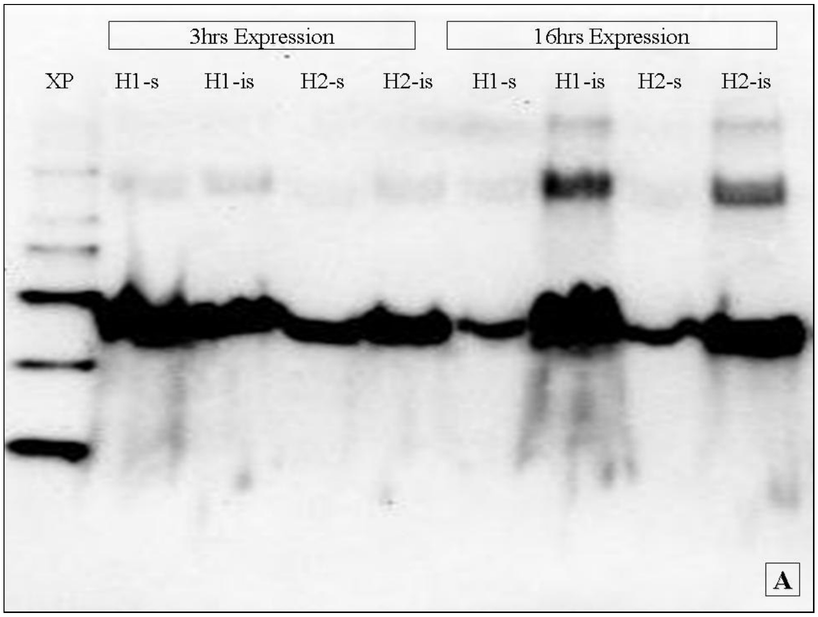

Expression for 3hrs
XP 15°C 22°C 30°C 37°C 42°C
A
Expression for 18hrs
XP 15°C 22°C 30°C 37°C 42°C
B
Expression for 3hrs
M12 15°C 22°C 30°C 37°C 42°C
C

Figure 4.7.7: detection for the effect of temperature on soluble expression of TVS 4041. Where, M12 stands for Mark 12 protein standard.(A: Western Blot, 3hrs harvest; B: Western Blot, 18hrs harvest; SDS-PAGE, 3hrs harvest)

4.7.8. Insolubility screening with increasing time of expression:

Soluble and insoluble fractions from H1 (pro-sequence, His tagged construct) and H2 (without pro-sequence, His tagged construct) were evaluate for 3hrs and 16hrs. It has been noticed that the H1 construct is better in overall soluble and insoluble expression then the H2 construct. Soluble fractions from H1 yielded more than H2 construct, for initial 3 hrs of expression. After 16 hrs, conversion of soluble protein in insoluble fraction has increased for both of the constructs. The tendency of soluble protein to become insoluble with time has proved to be more prone to H1 construct in comparison with H2 construct (figure: 4.7.8). Therefore, the soluble product from H1 and H2 constructs became similar in quantity after 16hrs. In understanding the ability of pro-sequence's **tendency** to express more is hidden in its 10 a.a sequence (ATEEFSVTPY). Since this sequence have more charged residues, therefore, overall tendency to get soluble yield increases according to prediction from section 4.3.8 (Wilkinson and Harrison 1991).

The segment of amino acid sequence in *C*-terminal tail of TVS4041 is predicted to contain transmembrane sequence (see section 4.4.1 for prediction) or a highly insoluble sequence (see section 4.3.8 for prediction). At increasing expression condition we therefore believed that TVS4041 tends to form insoluble aggregates. Hence, it is also impossible to purify in absence of detergent, as discussed in the purification part. These results provide strong evidence for the necessity to chopp off the *C*-terminal sequence in order to get a soluble and higher yield product.



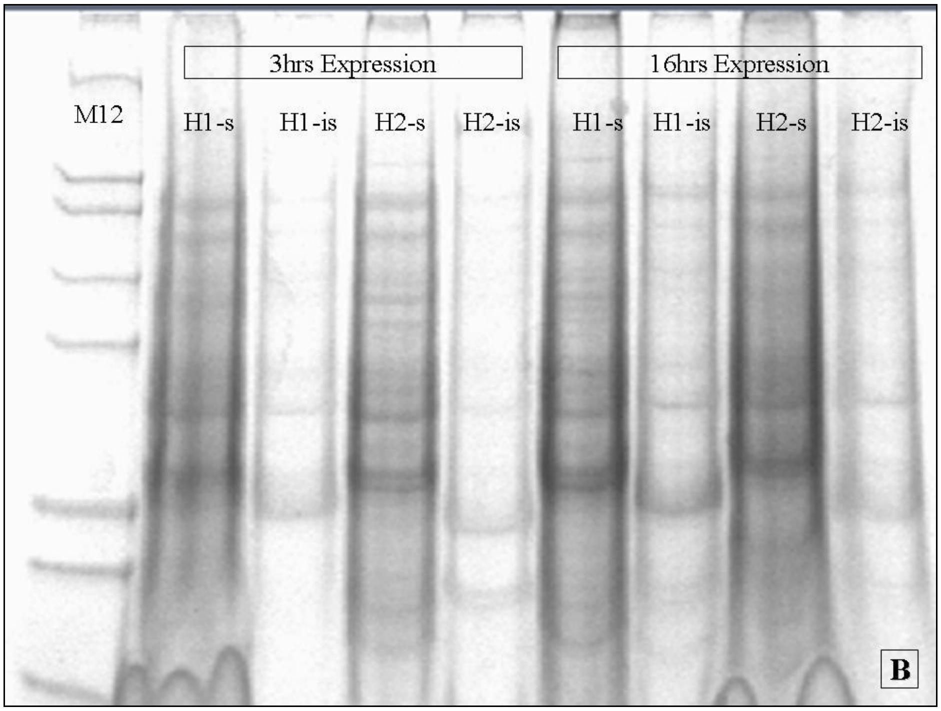


Figure 4.7.8: Detection of soluble and insoluble product. Where, M12 stands for Mark 12 protein standard, XP stands for XP magic marker. H1/H2-s= soluble fraction from H1/H2 construct, H1/H2-is= Insoluble fraction from H1/H2 construct. (A: Western Blot; B: SDS-PAGE)

4.7.9. Secretion of target protein into the media:

Although *E. coli* is generally not thought to be a good candidate for the secretion of cloned protein, but vigorous expression caused outer membrane permeabilization and leakage of the fusion protein into the culture medium (Paal, Heel et al. 2009).

In connection to the decreased soluble yield with increasing time (results from, section 4.7.4), an investigation was made to detect the possible leakage of target protein into the medium. When 10x concentrated medium from the 3hrs and 16hrs expressed culture were run on the gel, both of the constructs have shown the secretion capability into the medium (figure: 4.7.9). But similar to the fact that yielding capability of the H1 construct is higher than the H2 construct, the secreted amount is also in accordance with that pattern.

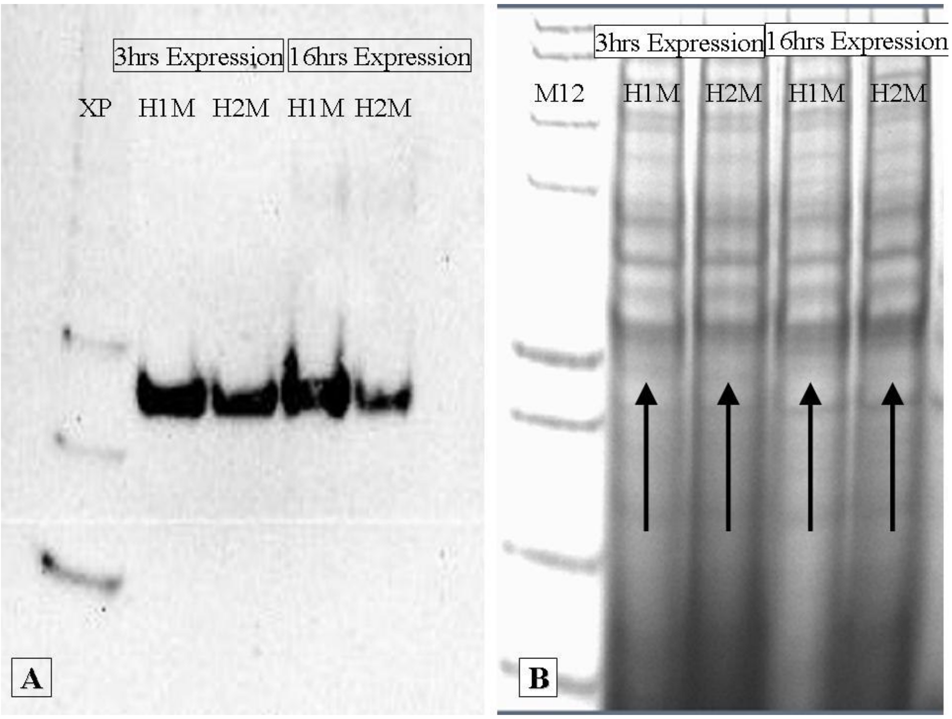


Figure 4.7.9: detection of secreted target protein (10x conc.) into the media. Where, M12 stands for Mark 12 protein standard, XP stand for Magic marker XP. (A: Western Blot; B: SDS-PAGE)

4.7.10. Effects of glucose on production of target protein:

Glucose is commonly known to stop the basal expression in case of toxic protein expression, but it is also predicted to increase the stability of the plasmid (Keevil, Spillane et al. 1987; Zhang, Taiming et al. 2003). Hence in this experiment it was suspected that overall expression would be improved by attaining the plasmid stability. On the other hand, presence of glucose can delay the expression process, consequently overnight stability of the target protein will improve.

Contrary to assumptions, obtained results from this experiment showed that the presence of glucose is not at all better for the expression in TOP10/pBADgIII expression system. In comparision from 3hrs to overnight expression, no considerable improvement occurs for soluble target protein production (figure: 4.7.10). Thus it can be concluded that the addition of glucose in any range above 0% can cause the *ara Operon system's* inability to be induced by arabinose in TOP10 cells, even after 20hrs.

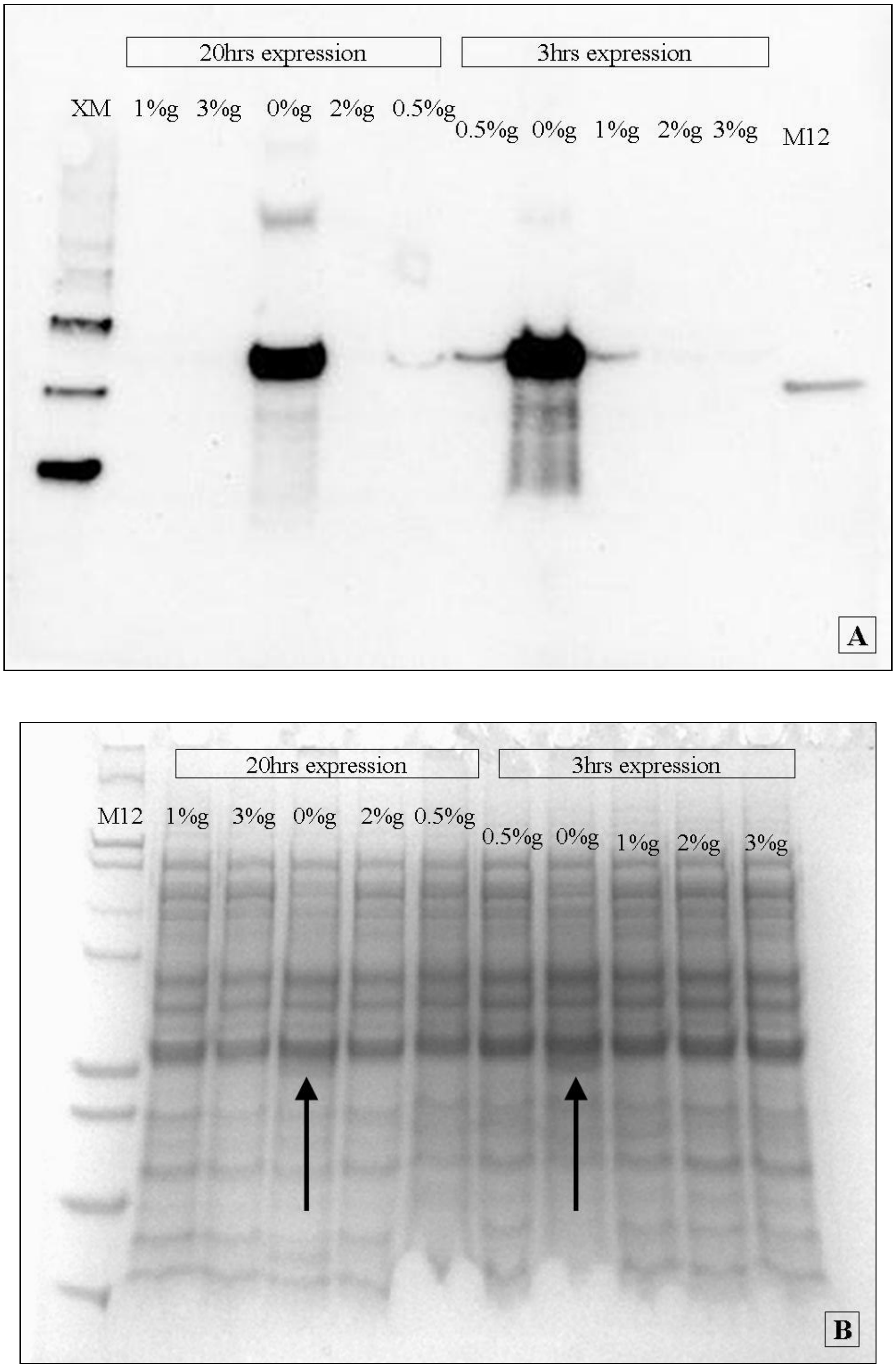


Figure 4.7.10: Effects of glucose on prolongated expression of the target protein. Here, % represents amount of glucose supplemented in 100ml media (g/100ml), M12 stands for Mark 12 protein standard, XP stands for Magic marker XP. (A: Western Blot; B: SDS-PAGE)

4.7.11. Ampicillin and Carbencillin for long period induction:

Plasmid selectivity of culture, grown with 100 µg/ml ampicillin at 37°C (at higher metabolic rate) could loss over the increasing time period. Plasmid vector harbours the β-Lactamase gene to survive in the presence of ampicillin, which is called the selectivity marker. This β-Lactamase gradually releases into the medium and at a certain density of cells, it can destroy all of the ampicillin from the medium. This results in a non selective culture medium, where tendency of plasmid loss appears in newly regenerated cells that do not carry the vector plasmid, hence unable to produce protein of interest and concluded with erroneous results. A strategy can be adapted to increase the concentration of ampicillin. An experiment was setup in order to see the effects of increasing ampicillin in medium and its effects for prolong induction of cell culture to fully benefit from a single batch of culture in the form of greater yield. Non-selectivity and plasmid-less culture at the late growth phase can be controlled by the addition of carbenicillin instead of vulnerable ampicillin. Since, carbencillin is more stable to β-Lactamase degradation and also stable at low pH (that arises usually at the end of late growth phase), and do not end up as a toxic substance upon degradation (Slocombe and Sutherland 1969; Basker, Comber et al. 1977; Pawelczyk, Zajac et al. 1981).

Moreover, (Taylor, Anderson et al. 1999) have described β–Lactam as a natural inhibitor of wide range of serine proteases (Wilmouth, Kassamally et al. 1999). Therefore it was thought that increasing-ampicillin (β–Lactam compound), can cause this serine type of protease (TVS4041) to be get stable against autolysis or any other *E. coli* based degradation. Based on this hypothesis, ampicillin and carbencillin were applied in media with varying concentration. Western blot analysis revealed the small molecular weight shift in Ampicilline 500 µg/l and Carbencillin 50 & 100 µg/l. This could be due to the inability of signal peptide cleavage that was 18 amino acid sequence. Expression band thickness is little higher from ampicilline 300-500 µg/l, but not in case of carbencillin (figure: 4.7.11 A & B). On the other hand TOP10 cells are in optimal density with ampicilline 300-400 µg/l, while with ampicillin 500 µg/l cell density has reduce significantly, representing the toxicness at that concentration. Cell density was also remained lower with both of the tested carbencillin concentrations (figure: 4.7.11.C).

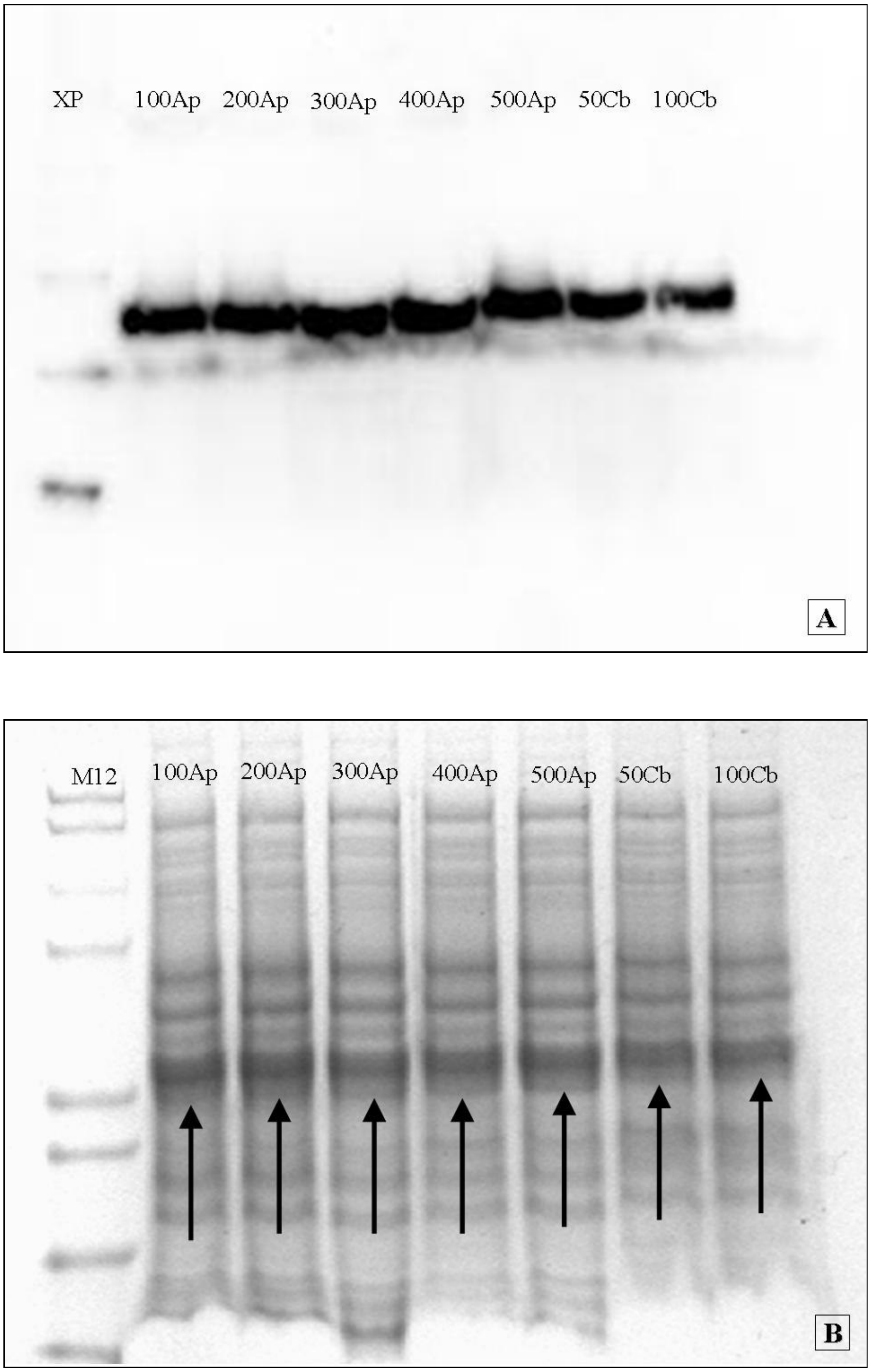


Figure 4.7.11: Effects of ampicilline and Carbencillin on prolongated expression of target protein. Where, **Ap** stands for ampicilline, **Cb** stands for carbencillin, **M12** stands for Mark 12 protein standard, **XP** stands for Magic marker XP. (A: Western Blot; B: SDS-PAGE)

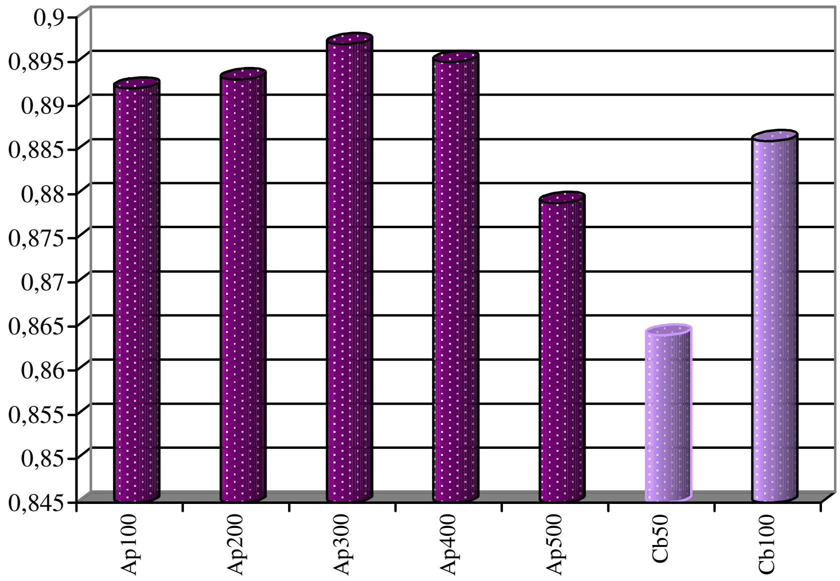


Figure 4.7.11.C: Effects of ampicilline and Carbencillin on TOP10 growth. While, Y-axis OD_{600}, X-axis describe the concentration of ampicillin and carbencillin.

4.7.12. Metal ions as additives for stable yield:

Metal ions have effects on serine proteases in three different ways, (i) they stabilize them against self digestion, (ii) they stabilize them in dilute solution and (iii) enhance the proteolytic activity of the proteases many folds. In some cases, they are essential for proteolytic activity (Mary, Achyuthan et al. 1988; Villalonga, Reyes et al. 2004). Several metal ions are reported to play important role in stability and catalytic efficiency of serine type of proteases like Ca^{2+} Zn^{2+}, Mg^{+}, Mn^{2+}, Co^{2+} and Cd^{2+} (Butler and Robins 1963; Gomez, Birnbaum et al. 1974; Kawasaki, Kurosu et al. 1986).

In addition, several metals and metal mixture are reported to increase the efficiency of protein expression (Studier 2005). Therefore, Ca^{2+}, Zn^{2+}, Mn^{2+}, $K^{+,}$ Mg^{+}, and Metal Mix (table 3.4.7) were included in the experimental evaluation for increased total yield of target protein. The general effect on TOP10 cells growth, with these metals supplementation into the media is plotted in graph against concentration of metal and observes cell density (OD_{600}) at harvest point (figure: 4.7.12, E).

The metal mixture was utilized in three different increasing concentrations, but increase above 1x (recommended concentration by (Studier 2005)) was not benificial since it results in declining target protein expression, but seems to enhance cell density slightly. As a whole if compared to condition, where no metal mix was supplied into the media (M0), then there were no desirable effect on total yield increment.

The protein under investigation does not show any increasing yield with increasing calcium chloride (1-10 mM) in the medium. Moreover, it can be seen that the target product seems to decrease with increasing calcium ($CaCl_2$), irrespective of overall good effects on the proliferation of cell density with increasing calcium ($CaCl_2$) (figure 4.7.12, E). However, later it was observed that $CaCl_2$ was causing milkyness in the 2xYT media that occluded the cell density measurments (OD_{600}, at harvest). Therefore based on the incorrect cell density, resuspension volumes are also altered from the actual cell density. Hence in case of effects of calcium chloride addition in media, detected declining target protein with increasing calcium chloride is dubious indication of pattren.

$ZnCl_2$ in any concentration above 1mM is deleterious to TOP10 cell and to the target protein production, as seen from SDS-PAGE, western blot and the cell density graph (figure 4.7.12; A,B,E). $MnCl_2$ is not very toxic to cell growth in 1mM concentration, but on increasing concentration it becomes toxic, as can be seen in the cell density graph. On the other hand, target protein production damaged severely by adding $MnCl_2$ above 1mM, for this expression system and for this particular protein (figure 4.7.12; C,D,E).

KCl has overall good effect on cell growth and target protein production from 1mM to 10mM. $MgCl_2$ has increasing effects for cell growth from 1-10mM, but target protein product is maximum with 1mM concentration. In this experiment, KCl and $MgCl_2$ (in 1mM conc.) are the only metals that increased the target protein yield in comparison to control, where no metal was added. $MgCl_2$ is also desirable for higher cell density, similarly KCl has also good effects on cell growth at any increasing concentration (figure 4.7.12; C,D,E).

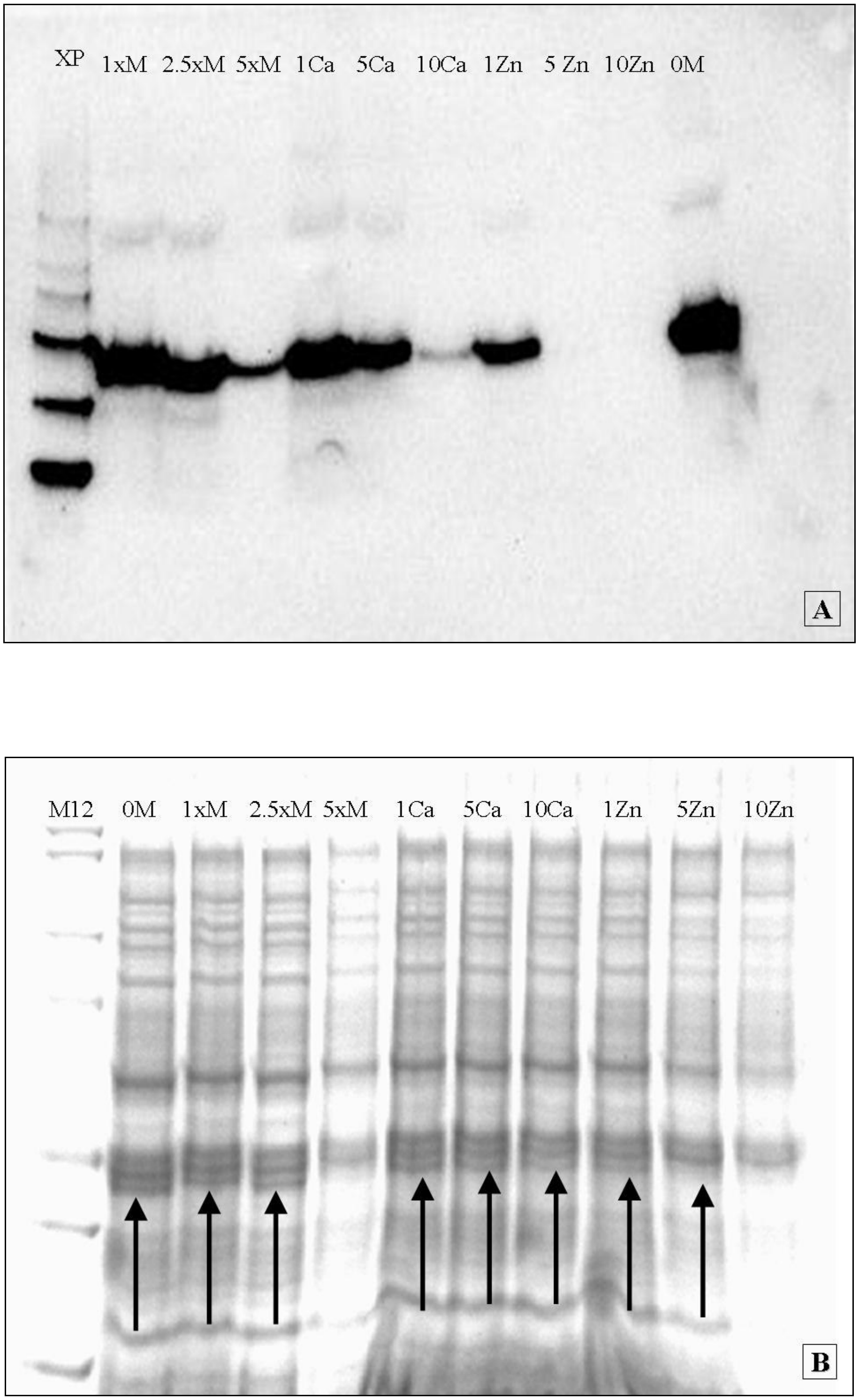

XP 1xM 2.5xM 5xM 1Ca 5Ca 10Ca 1Zn 5 Zn 10Zn 0M
A
M12 0M 1xM 2.5xM 5xM 1Ca 5Ca 10Ca 1Zn 5Zn 10Zn
B

Figure 4.7.12 (A, B): effects metals on total yield of target protein. Where, Ca and Zn denotes corresponding Chloride salts, taken in 1, 5, 10 mM concentration, 0M denotes no metal added (control), 1, 2.5, 5xM, denotes Metal Mix in increasing concentration. M12 stands for Mark 12 protein standard, XP stands for Magic marker XP. (A: Western Blot; B: SDS-PAGE)

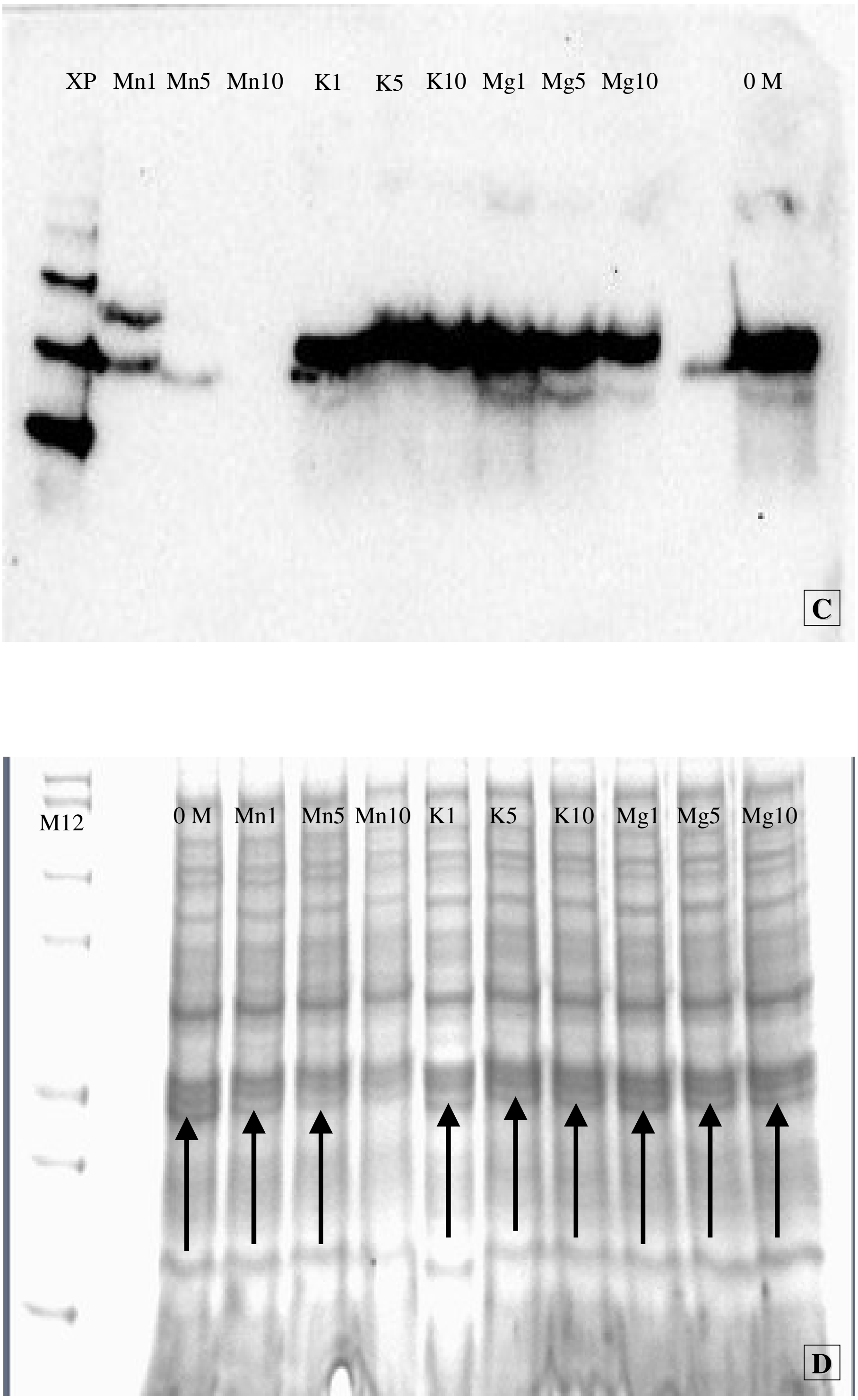


Figure 4.7.12 (C, D): effects metals on total yield of target protein. Where, Mn, K and Mg denotes corresponding Chloride salts, taken in 1, 5, 10 mM concentration. M12 stands for Mark 12 protein standard, XP stands for Magic marker XP. (C: Western Blot; D: SDS-PAGE)

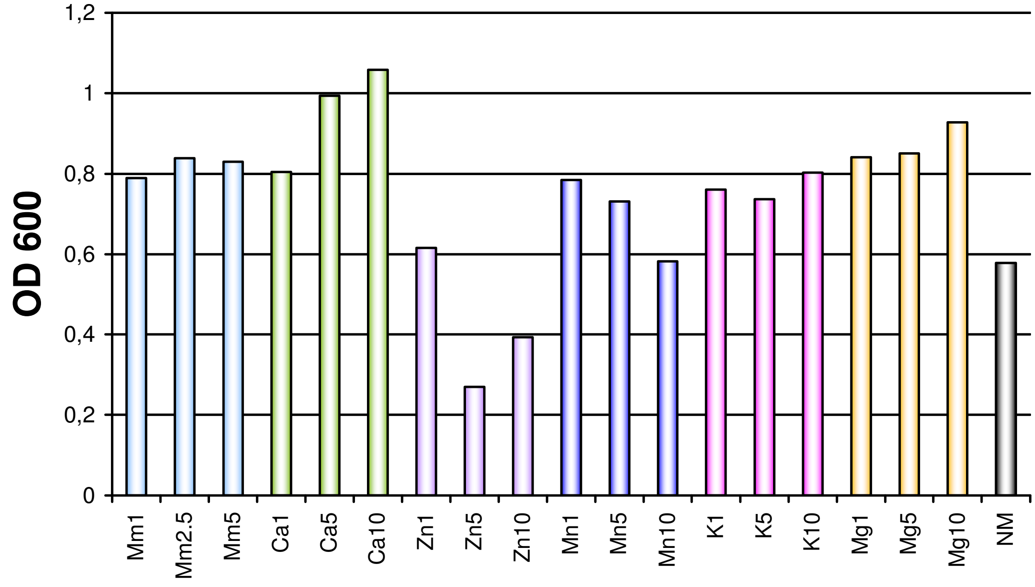


Figure 4.7.12.(E): effects of different metals on TOP10 growth. Y-axis denotes OD_{600} at harvest; X-axis denotes the concentration of metal salts of Ca, Zn, Mn, K, Mg used in these experiments (given in mM). Mm denotes Metal Mix use in 1x, 2.5x and 5x recommended concentration, NM denotes no metal addition (control).

4.7.13. Effects of aeration on solubility enhancement:

Aeration has drastic effects on the solubility of TVS4041. This phenomenon was observed by utilizing two different flask types in the experiment. One flask was normal with no inward structure, while the other flask type contains inward protrudes (sigma Aldrich), that works as propeller for increased aeration by breaking the swirl culture flow. Equal amount of cloned cultures were in both types of flask were grown in same conditions at 210 rpm, at 37ºC. Cultures grown in the serrated flask were better in growth and had proved to produce more soluble product than un-notched flasks (figure: 4.7.13). TVS4041 predicted to contains three disulphide bonds that assist in proper folding, solubility and stablity. With this experiment, we can conclud that this particular protein need higher aeration rate for oxidation process of cysteine to become cystine and help making the proper disulphide bridges in higher speed.

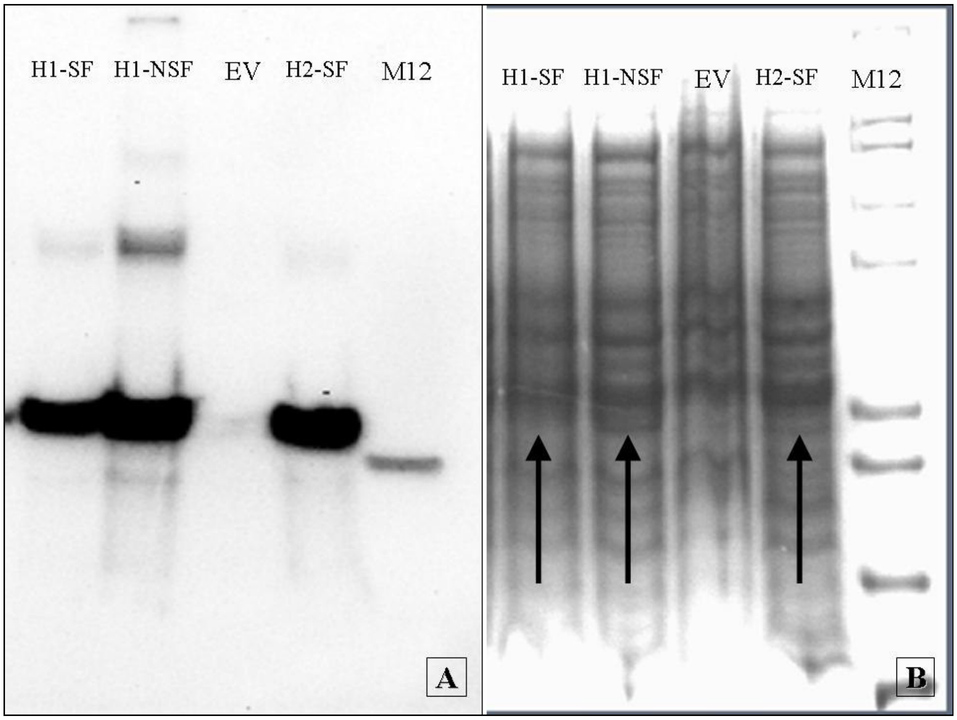


Figure 4.7.13: Effects of aeration on total soluble yield of target protein. Where, SF denotes serrated flasks, NSF denotes non serrated flasks. M12 stands for Mark 12 protein standard, XP stands for Magic marker XP.
(A: Western Blot; B: SDS-PAGE)

4.7.14. Comparison of all constructs soluble expression in modified conditions:

On comparing the soluble fraction of all constructs for soluble expression, then the order was N1>H1>>N2>>H2 (figure: 4.7.14; A, B). The constructs with pro sequence were better then the construct without the pro sequence in both native and His taged conditions. This phenomenon can be explained due to the property of pro sequence, which is predicted to be highly soluble in hydrophilic environment (section 4.3.8), hence capable of higher expression.

In comparision to the effects of the modified conditions to the unmodified conditions, a remarkable difference can be seen in the soluble expression increment when comparing figure 4.7.14 to figure 4.7.1. Hence it is concluded that 2xYT with lower salt and pH 6 can enhance the solubility of this target protein, provided that it should be express for shorter time period at 37°C with higher metabolic rate.

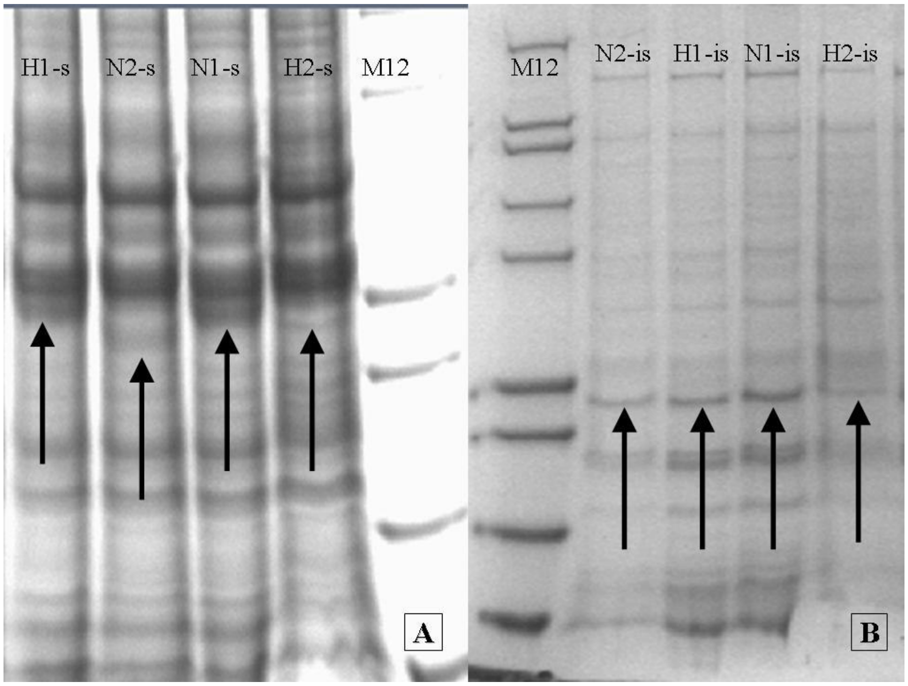


Figure 4.7.14: comparison of different constructs. Where, -s denotes soluble fraction, -is insoluble fraction, N1/H1 pro-sequence containing construct, with out His tag/ with His tag; N2/H2 with out pro-sequence containing construct, with out His tag/ with His tag. While, M12 stands for Mark 12 protein standard, XP stands for Magic marker XP. (A: Western Blot; B: SDS-PAGE)

4.8.0. TVS4041 characterization:

The theoretical pI values calculated for TVS4041 in mature form is 4.77 that is indicating this enzyme as an anionic type. This is also in accordance with the cold adapted feature of proteins in general (Siddiqui and Cavicchioli 2006). The homology model also confirms this assumption with a more negatively charged surface in comparision to the template from a thermostable species (figure: 4.3.13).

In attempts to characterize the TVS4041 for substrate specificity, all four substrates for trypsins, chymotrypsins and elastases (detail section 3.6.0) were employed separately into the periplasmic fraction of expressed culture. According to the results the only substrate that has reacted with the expressed periplasmic contents was **BAPNA,** indicating the trypsin activity of TVS4041. Among the four employed pH's 6, 7, 8, and 9, the activity only appears at pH 8 and 9. This is also in accordance with the results describe in section 4.7.6, where detection of TVS4041 through westrn blot was started to decline from pH 8 and 9 and a band of very low molecular range have been noticed from the sample grown at pH8 (degraded product shown with arrow in figure 4.7.6.A).

The activity of TVS4041 is also suspected to be salt dependent as seen from the figure 4.8.0, where cultures were grown over night with 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 percent salt in the media. Blue colored arrows represents the full length polypeptide, while in 1.5, 2.0, 2.5 percent salt a degraded form of this enzyme can be seen, represented in red arrows. This degraded product weights around 24 KD, that has also observed during the purification attempts, where lysis and running buffer were contained 500mM salt (figure: 4.10.2.C).

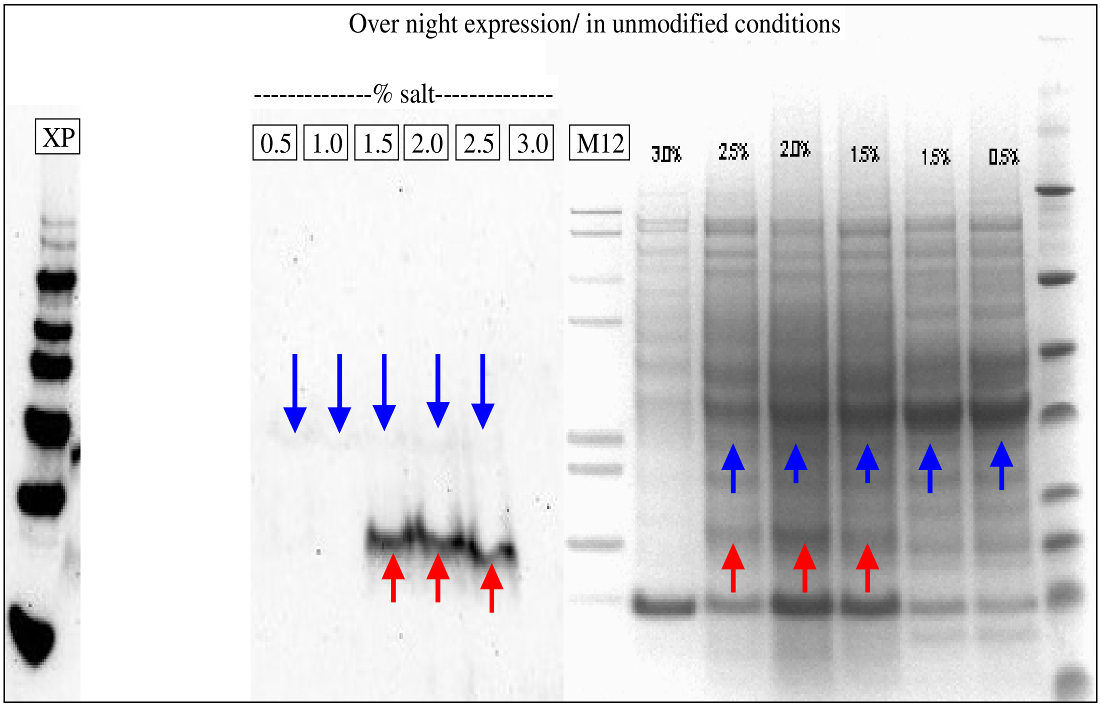

Over night expression/ in unmodified conditions
--------------% salt--------------
XP
0.5
1.0
1.5
2.0
2.5
3.0
M12
3.0%
2.5%
2.0%
1.5%
1.5%
0.5%

Figure 4.8.0: salt dependency for autolysis, shown with westrn blot and SDS-PAGE from over night express culture. Where, applied salt concentrations were taken in percent w/v concentration. M12 denotes Mark12, XP denotes Magic marker XP.

4.9.0. Soluble yield optimization for LexA, with different fusion tags:

Insolubility is serious bottleneck in heterologus protein production when using *E. coli* as a host organism. This phenomenon often leads to lower yield or inactive protein production (Nallamsetty, Austin et al. 2005). This problem has been in address of protein expression scientists, from the discovery of cloning and expression. Irrespective of other efforts (like media constituents and condition optimization) major efforts have been given to fusion of solubility enhancing proteins (fusion tag), that helps to properly fold the protein of interest (Esposito and Chatterjee 2006). We have used LexA in effort to compare the effects on yield and solubility of different fusion tags at at *N*-terminal of target protein.

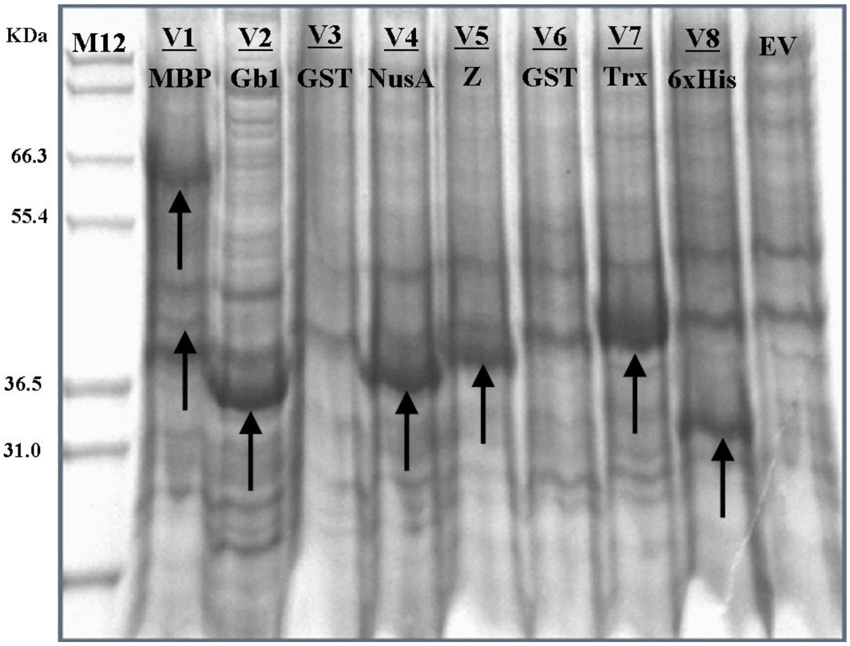


Figure 4.9.0: Expression of LexA cloned in to pDEST-TH1(V1), pDEST-TH3 (V2), pDEST-TH6 (V3), pDEST-TH7(V4), pDEST-TH10 (V5), pDEST 15 (V6), pDEST16 (V7), pDEST17(V8).

The following results were observed from the experiment:

- After three hours of expression (at 37 ºC, 240 rpm) there where very nominal amount of insoluble fraction, mostly seen for His tag construct. By resuspension of insoluble matrial in same volume as soluble fraction, no prominent bands were observed on the SDS-PAGE.

- SDS-PAGE analysis for soluble yield resulted in the following order for different fusion tags; Gb1>NusA>MBP=Trx=Z>6xHis and following order according to vectors; (pDEST-TH3>pDEST-TH7>pDEST-TH1=pDEST16=pDEST-TH10>pDEST17). Here pDEST17 is lowest copy number residing vector that has reported to contains about 15- 17 copies per cell (Sambrook, Fritsch et al. 1989). While other vectors are pUC origin that are known to have very higher copy number properties (500-700 per cell). But in expression of target protein performance of pDEST17 is not relatively too lower, as can be imagine from the low copy number.
- For unknown reasons, when GST taken as a fusion tag no protein can be detected as a visible over expressed band on SDS-PAGE. This is the case for both utilized vectors pDEST15 and pDEST-TH6. It might be possibile that the yield is much to low in order to see the strong bands on the SDS-PAGE. On an average, when looking at the results described in research work using various fusion tags, GST was found poor in soluble protein expression (Hammarstrom, Hellgren et al. 2002; Dyson, Shadbolt et al. 2004).
- In case of pDEST-TH1 (MBP as fusion tag) a full size product as well as a truncated product can be seen. The truncated product is of 40-50KD, that is around the weight of MBP solely. The similar weight band ~44KD has been observed in the work describe by (Niiranen, Espelid et al. 2007).
- Truncated product can also be seen with NusA as fusion partner. In case of pDEST-TH7 (NusA as fusion tag) only truncated product can be seen is in the size range of 36-40KD.

4.9.1. Discussion:

In most of the high throughput screening attempts with solubility tags, important issue is the method through which one can judge the degree of solubility. Observing the thickness of the expression band from the SDS-PAGE only by visual manner is not sufficient to judge the soluble expression. There are always higher possibilities of influenced decision and human miss judgment, especially when numerous numbers of experiments have to be evaluated side by side in highthrough put manner.

Another way to judge the solubility is by taking the estimation of soluble expression band intensity on SDS-PAGE, by using GelDoc software analysis. This estimation can also be misleading when expression band overlaps with host organism's indigenous protein. This phenomenon can also be seen in the LexA experiment where Trx-tagged LexA has overlapped one of the *E. coli* indigenous proteins (Figure 4.90, lane V7). In recent advancement of high through put expression studies, various numbers of authentic methods have been introduced to correctly evaluate the data statistically in 96- well plate format. The ELISA based method for solubility detection is an authentic way that produce the set of data best fitted for sensitive statistical evaluation with higher accuracy, reproducibility and with closer agreement to the SDS-PAGE based detection (Luan, Qiu et al. 2004).

Other approaches include the use of reporter proteins like green fluorescent protein (GFP), chloramphenicol acetyltransferase (CAT) for monitoring of soluble protein expression, but with disadvantage of extremely huge moiety co-expression (Maxwell, Mittermaier et al. 1999; Waldo, Standish et al. 1999). These method are now replaced by smaller protein *lacZα* (55 residue, detected by β-glactosidase activity) and S-tag (15 residues, detected by fluorescent assay) (Kelemen, Klink et al. 1999; Raines, McCormick et al. 2000; Wigley, Stidham et al. 2001; Eglen and Singh 2003).

One of the best options among all, is the activity determination for carrier protein, since sometime it has been observed that with the attachment of tag, target protein lose its activity in spite of its solubility. Even though activity measurment of all of the proteins are not available but for many of them where assays are reported it can directly confer the native confirmation of target protein.

When looking for the effects of solubility enhancement tags for a protein which originate from different source, then there is not one particular tag that always fit for all of the proteins from different sources and adapted to different climates. Although in highthrough put studies for solubility enhancement where several proteins are under studies, on average fewer tags always perform better, theses tags include MBP, NusA, and Trx (Dyson, Shadbolt et al. 2004; Busso, Delagoutte-Busso et al. 2005).

It can be suggest that solubility is the intrinsic property of protein to be study. Available solubility tags are proteins that have good capability of being adaptive to a broader range of buffer conditions and additives. Therefore by being solublize to a particular buffer condition, fusion protein force the attached accessory protein (target protein) to be suspend in that particular buffer conditions. Other hypothesis states the ability of fusion tags as a chaperone function in case of MBP (Kapust and Waugh 1999). Mutational evidences in fusion proteins present chaperone meganaets function for the solubility of carrier proteins (Fox, Kapust et al. 2001; Fox and Waugh 2003).

Maltose Binding Protein (MBP) and *N*-utilizing substance A (NusA) are very good solubility enhancing partners, but behave passively for the activity of protein that is the measurement of their true native confirmation (Nallamsetty and Waugh 2006). Mechanism of thioredoxin has some time works better than chaperon co-production, the main factor describe is its property to reduce the disulphide bonds (when bond formation is not required), that might be able to form abnormal structure and consequently aggregation. This has been observed in *E. coli* cytoplasm due to the relative oxidation environment in comparison to Mammalian cells (Yasukawa, Kanei-Ishii et al. 1995; Wong, Cai et al. 2004).

Sometimes when target proteins have intense tendency toward insolubility, fusion tags works as shield around them in the form of large micelle-like aggregates (Sachdev and Chirgwin 1999; Nomine, Ristriani et al. 2001). In this way carrier proteins remains as soluble aggregates and usually end up in inability to be purified.

4.10.0. Purification of TVS4041:

Two stretegies were attempted to capture the TVS4041 from crude extract by affinity chromatography techniques. These techneques involved benzamidine affinity for serine proteases on benzamidine (inhibitor) immobilized column and metal affinity for HisTagged proteins by immobilized metal ion chromatography (IMAC).

4.10.1. Benzamidine column:

The applied fractions from benzamidine column (GE Healthcare™, Bio-science AB), when eluted with 1M salt in a steep gradient resulted in three fractions of 1ml each (F3, F4 and F5). The eluted fractions had exhibted the BAPNA activity (Sigma Aldrich). Fraction F4 and flow-through (FT) both were showing the similar strength of BAPNA activity. This demonstrated the lower binding affinity of TVS4041 for benzamidine column in repoted conditions.

All of the eluted fractions F3, F4 and F5 were further concentrated (500µl from 3ml) and applied onto the gel filtration column (Superdex 200; GE Healthcare™). After a void volume a peak has appeared, all of the fractions obtained in this separation were again subjected for enzyme activity with BAPNA, resulted as fraction (F17) exhibited the BAPNA activity. The ultimate yield was much lower in that end part and no striking visible protein band was observed on SDS-PAGE, in the range of expected molecular weight.

Overall it was concluded that presence of target protein in flow through is due to the lower affinity of TVS4041 toward benzamidine column or either buffer conditions needed to be optimize. However, this phenomenon might be reduced by re-running of periplasmic fraction onto the column. Application of larger and concentrated volume was another recommendation to obtain the greater amount of target protein in eluted fraction. Another suspection was regarded with the presence of pro sequence in N1 construct, or the extra longer *C*-terminal that might hindered in between the binding site of benzamidine and protein.

Therefore, it was decideded from these results that another purification round must be run with H2 construct (His tagged without pro sequence) in a greater volume of concentrated periplasmic fraction. In the recommended conditions H2 construct (expresse in modified conditions for 3hrs, as describe in section 4.7), was re-run through benzamidine column, but no protein could be trapped, at that time.

4.10.2. HisTrap HP™ Column

Several trials made with crude extract and periplasmic fraction to purify TVS4041 on HisTrap HP™ column. The starting purification attempts were made with the periplasmic fractions from unmodified over night expression conditons, where soluble fraction contains very less amount of target protein, as describe in seaction 4.7.

In initial attempts periplasmic fractions were directly used to apply on HisTrap column, with results of no significant binding in the expected size range. These attempts were then realized to be vain, due to the presence of 2.5mM EDTA in the periplasmic fraction preparation buffer that was too high to strip the Ni^{++} from HisTrap column. Therefore a strteggy was designed to concentrate the periplasmic fraction to 1/8 of the volume and then increase the volume with running buffer to ½ of periplasmic volume.

A series of experiments were run on the OS1 (spheroplasmic fraction), OS2 (Periplasmic fraction), and concentrated medium from a same batch of cultures. It was seen from the OS1 and OS2 purification attempts that TVS4041 along with other HisTrap bounded proteins fall off in a single peak between ranges of 13-25% of elution buffer (25mM Tris HCl pH7.5, 10mM $CaCl_2$, 500mM Imidazole).

The strongest peak was observed in OS2 (Periplasmic fraction) while very small peak from OS1 fraction. Results obtained through the purification were run on SDS PAGE and were also cross check through His Tag western blot. On SDS-PAGE (Precise™ 4-20% protein gels) there was no strength, visible band detected in the range of expected intact polypeptide with in denaturing reducing conditions. While detected through Western Blot, very week purification strength was observed. Slight traces were also observe in the range of 60KD (can also be seen in western blots from expression experiments). This might represents dimer that could be formed due to the intra polypeptide cysteine bridges (during misfolding process), or can be a type of dimer due to the intra polypeptide stable secondary structure between complex C-terminal tail sequences. The eluted peack also showed an extremely rapid BAPNA activity that was observed from eluted fractions F10, F11, and F12.

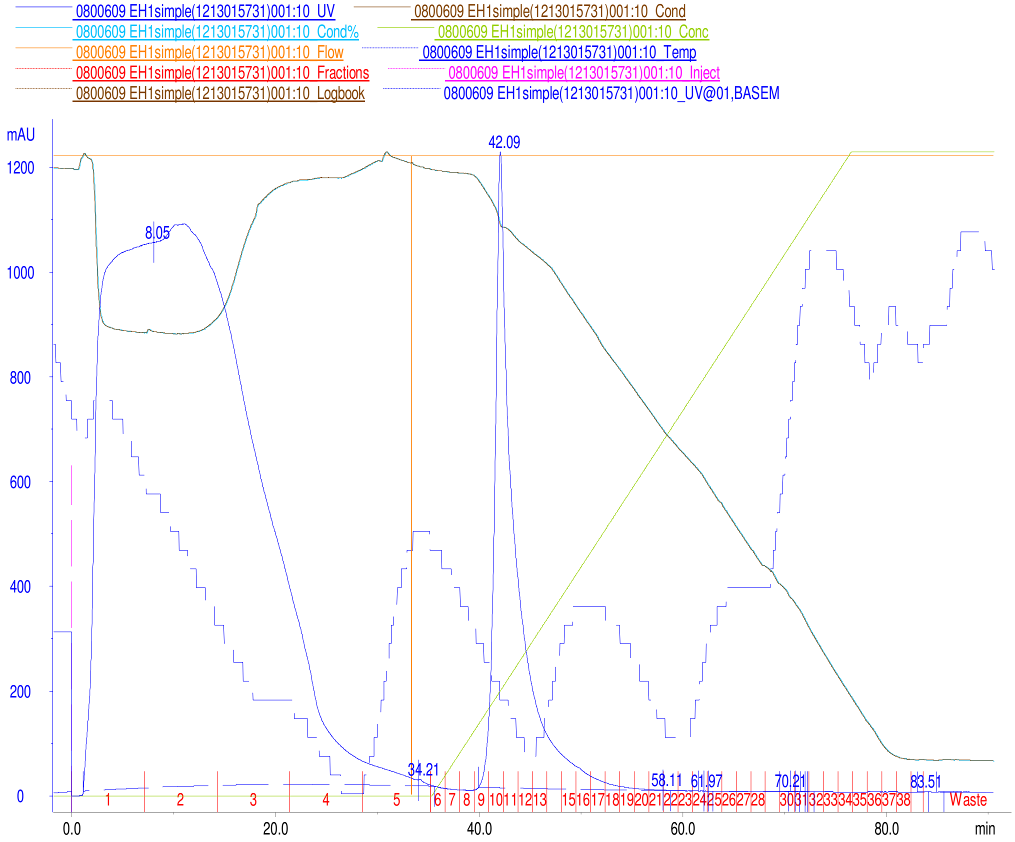


Figure 4.10.2 (A): chromatograph of purification from OS2 (periplasmic fraction).

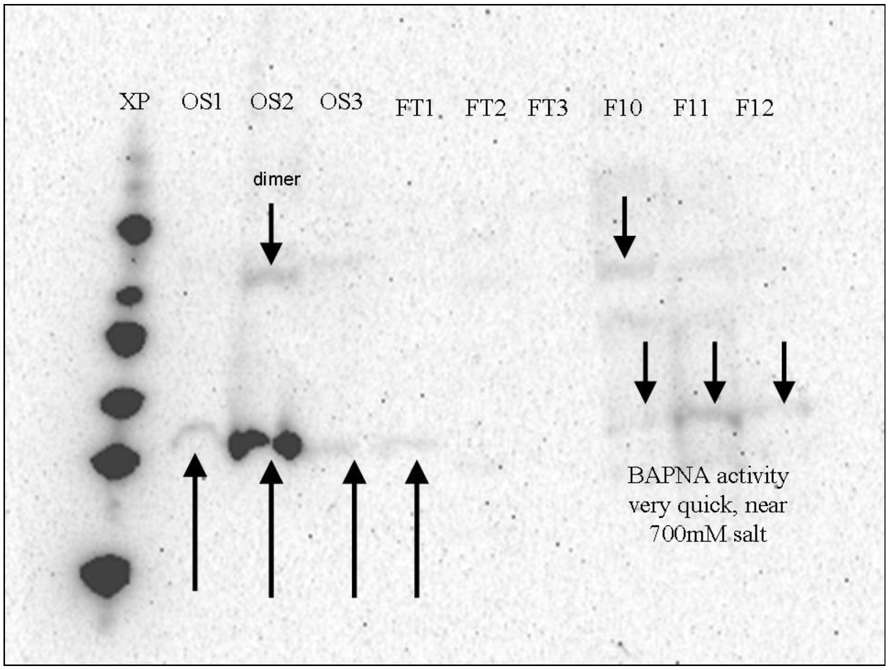


Figure 4.10.2 (B): Purification from OS2 (periplasmic fraction), where OS1 denotes Omotic shock solution1, OS2 periplasmic fraction, OS3 cellular fraction, FT are flow through fractions, F10,11 & 12 are eluted peake fractions corresponding to figure 4.10.2 (A).

Purification attempts from concentrated media resulted in a sharp peak around 40-50% elution buffer (25mM Tris HCl pH7.5, 10mM $CaCl_2$, 500mM Imidazole), while on SDS-PAGE no visible bands could be seen in the expected size range, probably due to the sedementation of protein that was observed as a deep blue color on the top of the wells.

Overall summery of theses results could be concluded as follows: binding affinity was not found much efficient when diluted sample were applied, running rate 0.5 ml/min or lower was found helpfull in efficient binding, Rerunning of the sample on the column with the help of peristelitic pump was found beneficial to recover more protein that was elude in flow through. Based on the results from bioinformatic analysis of TVS4041 (section 4.3.8), a transmembrane or highly hydrophobic segment is present in the *C*-terminal part of this protein that might form soluble aggregates and interfere in the purification process. Generally proteins from cold adapted sources contain hydrophobic patches on their surfaces that cause the difficulty in absence of hydrophobic elements solublelizing agents (such as detergents).

Keeping this in mind, a purification attempt was made with Triton X-100 in the running buffer (50mM Tris HCl pH 7.5, 750mM salt, 1% glycerol, 0.1% Triton X-100). The sample was re-ran on the column by peristaltic pump for 3 times and then eluted with ÄKTA FPLC, but the composition of running buffer interfeared with the UV absorption at 280nm, therefore no elution peak could be observed. Then all of the fractions were checked for the BAPNA activity that resulted in no detection. When detected, the samples on the western blot, presence of TVS4041, in applied fraction, flow through and eluted fractions were confirmed. TVS 4041 was seen in degraded form as shown with red arrows and in correct molecular weight size shown with blue arrows in figure 4.10.2 (C).

By looking into the results from westrn blot, TVS4041 was estimated to start eluting in 40% elution buffer (running buffer + 500mM Imidazole). But when crossed checked on SDS-PAGE, not a single band could be detected from the eluted fractions. This suggests that this protein might goes into the self degradation, or it precipitated from the eluted fractions. In case there was a chance of degradation, reversible/ irreversible protease inhibitors need to be used to avoid any self-digestion. While for avoiding the precipitation repeated freez and thaw should not be in practice, specially when protein is from cold adapted source, since proteins from cold adapted sources are more vulnerable to cold denaturation (Siddiqui and Cavicchioli 2006).

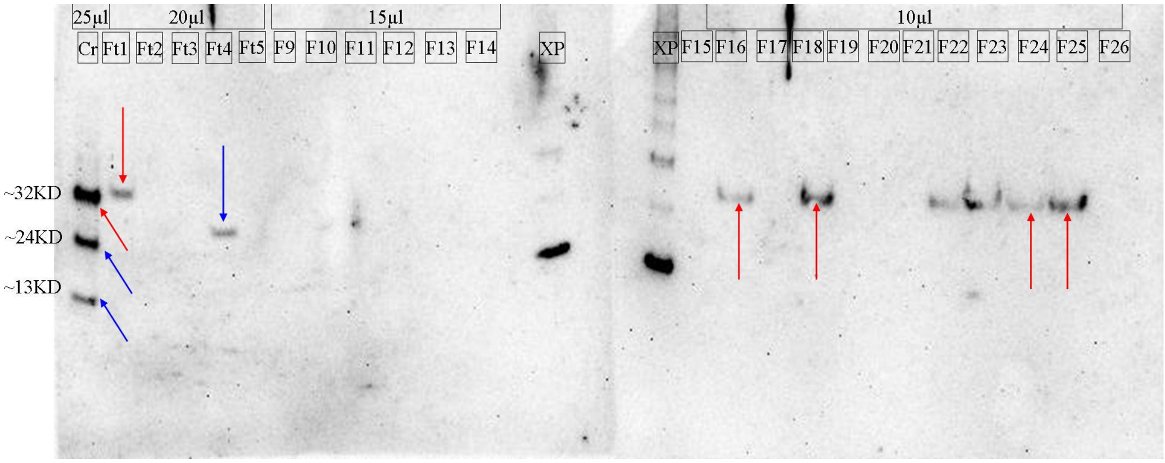


Figure 4.10.2 (C): purification attempts with Triton X100, **Cr** denotes crude extract, **Ft** denotes flow through fractions and **F** denots eluted fractions. Blue arrow denotes lysed product of TVS4041 in crude extract and in Ft4, red arrows denotes complete polpeptide in crude extract, Ft1 and eluted fractions F16-F25.

Conclusion

&

Future prospective

The series of experimental conditions evaluation for the TVS4041, resulted in modifies salt, pH and harvesting time conditions, where most of the protein now remains soluble. These experimentally evaluated condition, can be altered during purification hence can cause the aggregation or precipitation. Therefore it is recommended that target protein TVS4041 needed to be further clone without *C*-terminal domain (suspected, major insolubility domain) to avoid any possible clump formation during purification. We hope, this domain clipping will be the ultimate solution for increase soluble expression and higher yield purification for TVS4041.

Secreted properties of target protein can be exploited to produce a continuous flow system for the isolation of protein directly from medium. This method can be advantagious as by this mean we can avoide the damaging side effects like, heat shearing and chemical mediated inactivation of enzymes (Fish and Lilly 1984). Glycine and other chemicals are known to have effect on Periplasmic proteins to be released into the medium (Yu, Aristidou et al. 1991). An experimental evaluation is needed to see the effects of these applied factors for secretion of TVS4041.

The problem in the increasing insolubility with time can also be overcome by exporting the protein out of cell, which is also a good option to minimize the purification steps and cost. Most of the industrially applicable enzymes produce by this way in greater yield and further recover with low cost by just filtering the media broth and concentrate by evaporation or by ultra-filtration. The resulting concentrated material then precipitated by adding inorganic solvent (Vromen, A, J, 1997).

From the expression and purification studies of TVS4041, an initial characterization is now predicted for this particular target. According to these assumptions TVS4041 is less stable and autolysing at pH 7-9 and salt concentration 1-3 percent (w/v). Therefore, it is concluded that for purification lower pH (6) and no salt will be significantly improve the yield of stable TVS4041.

In the field of protein expression several attempts have been made and are still in struggling to look for the best fusion partner for the solubility of insoluble proteins. Here we have tried to express LexA from cold adapted bacterial species, in different expression vector and with different fusion tags in connection to the *N*- terminal of target protein. Findings concluded in the pattern where Gb1>NusA>Z=Trx>MBP>6xHis. These results are somewhat different then similar work from the same organism but with five different proteins, two different expression host and three different temperature effects. The overall average pattern from five different proteins and two different temperature was found as MBP>NusA>>Gb-1>Trx>GST>Z>6xHis (Niiranen, Espelid et al. 2007). The contradiction of results simply explains the variation of protein nature within the genome and broadly to the cold adaptive group.

Irrespective of beneficial use of solubility tags in expression, utilization of solubility tags some time brings a set of problems in addition to increasing the yield and solubility (Esposito and Chatterjee 2006). These problem accounts for:

(i) It takes extra efforts of cell protein synthesis machinery in the expression of large solubility tags like NusA (56 KD), MBP (45 KD), GST (27 KD), Trx (14.3 KD) etc.
(ii) Some of the fusion tags end up with inability of cleavage, due to unavailable linker region.
(iii) some proteins precipitate after the cleavage of tags
(iv) some proteins lose the activity even though they remain soluble in company with fusion tags (Sachdev and Chirgwin 1998; Sachdev and Chirgwin 1999). It could be due to the conformational change in the active site in the presence of fusion tag, or could be due to the unavailability of the active site.

Hence, it is recommended that if fusion tags should be avoided to use. But if it is essential to use them, then each and every target should be evaluated separately with different fusion tags. Since, all of the proteins do not behave similarly, though they are from same climate and same source of organism.

Based on the experienced from TVS4041 insolubility analysis (table 4.3.8), it is strongly recommend that all of the selected target should be essentially analyzed thoroughly, for any unusual complex domain that might cause problem in solubility and in turn hinder in future progress.

References

Abraham, C. R. and H. Potter (1989). "The protease inhibitor, alpha 1-antichymotrypsin, is a component of the brain amyloid deposits in normal aging and Alzheimer's disease." Ann Med **21**(2): 77-81.

Abrahmsen, L., T. Moks, et al. (1986). "Secretion of heterologous gene products to the culture medium of Escherichia coli." Nucleic Acids Res **14**(18): 7487-500.

Abramowitz, N., I. Schechter, et al. (1967). "On the size of the active site in proteases. II. Carboxypeptidase-A." Biochem Biophys Res Commun **29**(6): 862-7.

Adekoya, O. A., R. Helland, et al. (2006). "Comparative sequence and structure analysis reveal features of cold adaptation of an enzyme in the thermolysin family." Proteins **62**(2): 435-49.

Altschul, S. F., W. Gish, et al. (1990). "Basic local alignment search tool." J Mol Biol **215**(3): 403-10.

Altschul, S. F. and D. J. Lipman (1990). "Protein database searches for multiple alignments." Proc Natl Acad Sci U S A **87**(14): 5509-13.

Altschul, S. F., J. C. Wootton, et al. (2005). "Protein database searches using compositionally adjusted substitution matrices." Febs J **272**(20): 5101-9.

Alvarez, M., J. P. Zeelen, et al. (1998). "Triose-phosphate isomerase (TIM) of the psychrophilic bacterium Vibrio marinus. Kinetic and structural properties." J Biol Chem **273**(4): 2199-206.

Andersen, C. L., A. Matthey-Dupraz, et al. (1997). "A new Escherichia coli gene, dsbG, encodes a periplasmic protein involved in disulphide bond formation, required for recycling DsbA/DsbB and DsbC redox proteins." Mol Microbiol **26**(1): 121-32.

Anson, D. S. and K. R. Dunning (2005). "Codon-optimized reading frames facilitate high-level expression of the HIV-1 minor proteins." Mol Biotechnol **31**(1): 85-8.

Araiza Orozco, L. M., E. E. Avila Muro, et al. (1997). "Entamoeba histolytica: role of surface proteases on its virulence." Arch Med Res **28 Spec No**: 175-7.

Aslanidis, C. and P. J. de Jong (1990). "Ligation-independent cloning of PCR products (LIC-PCR)." Nucleic Acids Res **18**(20): 6069-74.

Baneyx, F. (1999). "Recombinant protein expression in Escherichia coli." Curr Opin Biotechnol **10**(5): 411-21.

Baneyx, F. and M. Mujacic (2002). E. coliGene Expression Protocols:Volume

Baneyx, F. and M. Mujacic (2004). "Recombinant protein folding and misfolding in Escherichia coli." Nat Biotechnol **22**(11): 1399-408.

Barbas, C. F., 3rd, A. S. Kang, et al. (1991). "Assembly of combinatorial antibody libraries on phage surfaces: the gene III site." Proc Natl Acad Sci U S A **88**(18): 7978-82.

Barnes, L. M. and A. J. Dickson (2006). "Mammalian cell factories for efficient and stable protein expression." Curr Opin Biotechnol **17**(4): 381-6.

Barrett, A. J. (1995). "Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). Enzyme nomenclature. Recommendations 1992. Supplement 2: corrections and additions (1994)." Eur J Biochem **232**(1): 1-6.

Barrett, A. J. (1996). "Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). Enzyme nomenclature. Recommendations 1992. Supplement 3: corrections and additions (1995)." Eur J Biochem **237**(1): 1-5.

Barrett, A. J. (1997). "Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). Enzyme Nomenclature.

Recommendations 1992. Supplement 4: corrections and additions (1997)." Eur J Biochem **250**(1): 1-6.

Barrett, A. J., M. A. Brown, et al. (1995). "Thimet oligopeptidase and oligopeptidase M or neurolysin." Methods Enzymol **248**: 529-56.

Barrett, A. J. and J. K. McDonald (1985). "Nomenclature: a possible solution to the 'peptidase anomaly'." Biochem J **231**(3): 807.

Barrett, A. J. and N. D. Rawlings (1991). "Types and families of endopeptidases." Biochem Soc Trans **19**(3): 707-15.

Barrett, A. J. and N. D. Rawlings (2007). "'Species' of peptidases." Biol Chem **388**(11): 1151-7.

Barrett, A. J., N. D. Rawlings, et al. (2001). "The MEROPS database as a protease information system." J Struct Biol **134**(2-3): 95-102.

Barrett, A. J., Rawlings, N.D. and Woessner, J.F. (1998). Handbook of Proteolytic Enzymes. London, Academic Press.

Bartel, B., I. Wunning, et al. (1990). "The recognition component of the N-end rule pathway." Embo J **9**(10): 3179-89.

Basker, M. J., K. R. Comber, et al. (1977). "Carfecillin: antibacterial activity in vitro and in vivo." Chemotherapy **23**(6): 424-35.

Baur, X. and G. Fruhmann (1979). "Allergic reactions, including asthma, to the pineapple protease bromelain following occupational exposure." Clin Allergy **9**(5): 443-50.

Beher, D. and S. L. Graham (2005). "Protease inhibitors as potential disease-modifying therapeutics for Alzheimer's disease." Expert Opin Investig Drugs **14**(11): 1385-409.

Bendtsen, J. D., L. Kiemer, et al. (2005). "Non-classical protein secretion in bacteria." BMC Microbiol **5**: 58.

Berger, A. S., I. (1970). "Mapping the active site of papain with the aid of peptide substrates and inhibitors." Philos. Trans. R. Soc. London, Ser. B:Biol. Sci. **257**: 249-264.

Bergmann, M. and W. F. Ross (1936). "On proteolytic enzymes, X. The enzymes of papain and their activation." Biol. Chem. **114**(2169): 717-26.

Bieker, K. L. and T. J. Silhavy (1990). "The genetics of protein secretion in E. coli." Trends Genet **6**(10): 329-34.

Blaber, M. (1998, Spring). "Molecular Biology and Biotechnology." Retrieved 17.02.2009, from www.mikeblaber.org/.../lect16/IMG00001.GIF.

Bode, W., E. Meyer, Jr., et al. (1989). "Human leukocyte and porcine pancreatic elastase: X-ray crystal structures, mechanism, substrate specificity, and mechanism-based inhibitors." Biochemistry **28**(5): 1951-63.

Bolivar, F., M. C. Betlach, et al. (1977). "Origin of replication of pBR345 plasmid DNA." Proc Natl Acad Sci U S A **74**(12): 5265-9.

Bolivar, F., R. L. Rodriguez, et al. (1977). "Construction and characterization of new cloning vehicles. I. Ampicillin-resistant derivatives of the plasmid pMB9." Gene **2**(2): 75-93.

Bolivar, F., R. L. Rodriguez, et al. (1977). "Construction and characterization of new cloning vehicles. II. A multipurpose cloning system." Gene **2**(2): 95-113.

Boy, S., C. Seif, et al. (2008). "[Botulinum toxin in the treatment of benign prostatic hyperplasia : An overview.]." Urologe A.

Branden, C. and J. Tooze (1999). Folding and Flexiblity. NY, Gardland Publishing.

Braud, S., M. Moutiez, et al. (2005). "Dual expression system suitable for high-throughput fluorescence-based screening and production of soluble proteins." J Proteome Res **4**(6): 2137-47.

Busso, D., B. Delagoutte-Busso, et al. (2005). "Construction of a set Gateway-based destination vectors for high-throughput cloning and expression screening in Escherichia coli." Anal Biochem **343**(2): 313-21.

Butler, J. A. and A. B. Robins (1963). "Effects of Certain Metal Salts on the Inactivation of Solid Trypsin by Ionizing Radiation." Radiat Res **19**: 582-92.

Cabrita, L. D. and S. P. Bottomley (2004). "Protein expression and refolding--a practical guide to getting the most out of inclusion bodies." Biotechnol Annu Rev **10**: 31-50.

Campbell, N. and J. Reece (2006). Biologie, Computer Press.

Carter, D. B., E. Dunn, et al. (2008). "Changes in gamma-secretase activity and specificity caused by the introduction of consensus aspartyl protease active motif in Presenilin 1." Mol Neurodegener **3**: 6.

Caspers, P., M. Stieger, et al. (1994). "Overproduction of bacterial chaperones improves the solubility of recombinant protein tyrosine kinases in Escherichia coli." Cell Mol Biol (Noisy-le-grand) **40**(5): 635-44.

Cassland, P., S. Larsson, et al. (2004). "Heterologous expression of barley and wheat oxalate oxidase in an E. coli trxB gor double mutant." J Biotechnol **109**(1-2): 53-62.

Chang, A. C. and S. N. Cohen (1978). "Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid." J Bacteriol **134**(3): 1141-56.

Chartrain, M. and L. Chu (2008). "Development and production of commercial therapeutic monoclonal antibodies in Mammalian cell expression systems: an overview of the current upstream technologies." Curr Pharm Biotechnol **9**(6): 447-67.

Chen, H., M. Bjerknes, et al. (1994). "Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of Escherichia coli mRNAs." Nucleic Acids Res **22**(23): 4953-7.

Chen, J., J. L. Song, et al. (1999). "Chaperone activity of DsbC." J Biol Chem **274**(28): 19601-5.

Chou, C. H., A. A. Aristidou, et al. (1995). "Characterization of a pH-inducible promoter system for high-level expression of recombinant proteins in Escherichia coli." Biotechnol Bioeng **47**(2): 186-92.

Colquhoun, D. J., K. Alvheim, et al. (2002). "Relevance of incubation temperature for Vibrio salmonicida vaccine production." J Appl Microbiol **92**(6): 1087-96.

Colquhoun, D. J. and H. Sorum (2001). "Temperature dependent siderophore production in Vibrio salmonicida." Microb Pathog **31**(5): 213-9.

Curry, S., N. Roque-Rosell, et al. (2007). "Foot-and-mouth disease virus 3C protease: recent structural and functional insights into an antiviral target." Int J Biochem Cell Biol **39**(1): 1-6.

Cushman, D. W. and H. S. Cheung (1971). "Spectrophotometric assay and properties of the angiotensin-converting enzyme of rabbit lung." Biochem Pharmacol **20**(7): 1637-48.

D'Amico, S., T. Collins, et al. (2006). "Psychrophilic microorganisms: challenges for life." EMBO Rep **7**(4): 385-9.

Das, S. and P. Mukhopadhyay (1994). "Protease inhibitors in chemoprevention of cancer. An overview." Acta Oncol **33**(8): 859-65.

de Backer, M., S. McSweeney, et al. (2002). "The 1.9 A crystal structure of heat-labile shrimp alkaline phosphatase." J Mol Biol **318**(5): 1265-74.

de las Rivas, B., J. A. Curiel, et al. (2007). "Expression vectors for enzyme restriction- and ligation-independent cloning for producing recombinant His-fusion proteins." Biotechnol Prog **23**(3): 680-6.

de Marco, A., E. Deuerling, et al. (2007). "Chaperone-based procedure to increase yields of soluble recombinant proteins produced in E. coli." BMC Biotechnol **7**: 32.

De Nanteuil, G., B. Portevin, et al. (2001). "Disease-modifying anti-osteoarthritic drugs: current therapies and new prospects around protease inhibition." Farmaco **56**(1-2): 107-12.

DeClerck, Y. A. and S. Imren (1994). "Protease inhibitors: role and potential therapeutic use in human cancer." Eur J Cancer **30A**(14): 2170-80.

Di Cera, E. (2008). "Engineering protease specificity made simple, but not simpler." Nat Chem Biol **4**(5): 270-1.

Doebber, T. W., A. R. Divor, et al. (1978). "Identification of a tripeptidyl aminopeptidase in the anterior pituitary gland: effect on the chemical and biological properties of rat and bovine growth hormones." Endocrinology **103**(5): 1794-804.

Dubendorff, J. W. and F. W. Studier (1991). "Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor." J Mol Biol **219**(1): 45-59.

Dyson, M. R., S. P. Shadbolt, et al. (2004). "Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression." BMC Biotechnol **4**: 32.

Eglen, R. M. and R. Singh (2003). "Beta galactosidase enzyme fragment complementation as a novel technology for high throughput screening." Comb Chem High Throughput Screen **6**(4): 381-7.

Emanuelsson, O., S. Brunak, et al. (2007). "Locating proteins in the cell using TargetP, SignalP and related tools." Nat Protoc **2**(4): 953-71.

Endo, S., Y. Tomimoto, et al. (2006). "Effects of E. coli chaperones on the solubility of human receptors in an in vitro expression system." Mol Biotechnol **33**(3): 199-209.

Engler, C., R. Kandzia, et al. (2008). "A one pot, one step, precision cloning method with high throughput capability." PLoS ONE **3**(11): e3647.

Esposito, D. and D. K. Chatterjee (2006). "Enhancement of soluble protein expression through the use of fusion tags." Curr Opin Biotechnol **17**(4): 353-8.

Evans, P. A., C. M. Dobson, et al. (1987). "Proline isomerism in staphylococcal nuclease characterized by NMR and site-directed mutagenesis." Nature **329**(6136): 266-8.

Fehlhammer, H. and W. Bode (1975). "The refined crystal structure of bovine beta-trypsin at 1.8 A resolution. I. Crystallization, data collection and application of patterson search technique." J Mol Biol **98**(4): 683-92.

Feinbaum, R. (2001). "Introduction to plasmid biology." Curr Protoc Mol Biol **Chapter 1**: Unit1 5.

Fish, M. and M. Lilly (1984). "The Interactions Between Fermentation and Protein Recovery." nature biotechnology **2**: 623-627.

Fox, J. D., R. B. Kapust, et al. (2001). "Single amino acid substitutions on the surface of Escherichia coli maltose-binding protein can have a profound impact on the solubility of fusion proteins." Protein Sci **10**(3): 622-30.

Fox, J. D. and D. S. Waugh (2003). "Maltose-binding protein as a solubility enhancer." Methods Mol Biol **205**: 99-117.

Friehs, K. (2004). "Plasmid copy number and plasmid stability." Adv Biochem Eng Biotechnol **86**: 47-82.

Froderberg, L., E. Houben, et al. (2003). "Versatility of inner membrane protein biogenesis in Escherichia coli." Mol Microbiol **47**(4): 1015-27.

Fu, W., J. Lin, et al. (2008). "Expression of a hemA Gene from Agrobacterium radiobacter in a Rare Codon Optimizing Escherichia coli for Improving 5-aminolevulinate Production." Appl Biochem Biotechnol.

Fuss, C. N. and K. O. Godwin (1975). "A comparison of the uptake of [75Se]selenite, [75Se]selenomethionine and [35S]methionine by tissues of ewes and lambs." Aust J Biol Sci **28**(3): 239-49.

Gasteiger, E., C. Hoogland, et al. (2005). Protein Identification and Analysis Tools on the ExPASy Server, Humana Press

GE, h. (2006) "Benzamidine sephrose 6B." **Volume**, DOI:

Goel, A., D. Colcher, et al. (2000). "Relative position of the hexahistidine tag effects binding properties of a tumor-associated single-chain Fv construct." Biochim Biophys Acta **1523**(1): 13-20.

Goguen, J. D., N. P. Hoe, et al. (1995). "Proteases and bacterial virulence: a view from the trenches." Infect Agents Dis **4**(1): 47-54.

Goldberg, B. and R. B. Stricker (1996). "HIV protease and the pathogenesis of AIDS." Res Virol **147**(6): 375-9.

Golovanov, A. P., G. M. Hautbergue, et al. (2004). "A simple method for improving protein solubility and long-term stability." J Am Chem Soc **126**(29): 8933-9.

Gomez, J. E., E. R. Birnbaum, et al. (1974). "The metal ion acceleration of the conversion of trypsinogen to trypsin. Lanthanide ions as calcium ion substitutes." Biochemistry **13**(18): 3745-50.

Gonda, D. K., A. Bachmair, et al. (1989). "Universality and structure of the N-end rule." J Biol Chem **264**(28): 16700-12.

Grassmann, W. and H. Dyckerhoff (1928). "Über die Proteinase und die Polypeptidase der Hefe. 13. Abhandlung über Pflanzenproteasen in der von R. Willstätter und Mitarbeitern begonnenen Untersuchungsreihe. ." Hoppe-Seyler's Z. Physiol. Chem. **179**: 41-78.

Greene, J. J. (2004). "Host cell compatibility in protein expression." Methods Mol Biol **267**: 3-14.

Grunnet, I. and J. Knudsen (1983). "Medium-chain fatty acid synthesis by goat mammary-gland fatty acid synthetase. The effect of limited proteolysis." Biochem J **209**(1): 215-22.

Grzesiak, A., R. Helland, et al. (2000). "Substitutions at the P(1) position in BPTI strongly affect the association energy with serine proteinases." J Mol Biol **301**(1): 205-17.

Guruprasad, K., B. V. Reddy, et al. (1990). "Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence." Protein Eng **4**(2): 155-61.

Hall, T. (1997). from http://www.mbio.ncsu.edu/BioEdit/bioedit.html.

Hammarstrom, M., N. Hellgren, et al. (2002). "Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli." Protein Sci **11**(2): 313-21.

Han, K. G., S. S. Lee, et al. (1999). "Soluble expression of cloned phage K11 RNA polymerase gene in Escherichia coli at a low temperature." Protein Expr Purif **16**(1): 103-8.

Hart, D. J. and F. Tarendeau (2006). "Combinatorial library approaches for improving soluble protein expression in Escherichia coli." Acta Crystallogr D Biol Crystallogr **62**(Pt 1): 19-26.

Hartley, B. S. (1960). "Proteolytic enzymes." Annu Rev Biochem **29**: 45-72.

Hasegawa, J. (1960). "Exopeptidases of the human skin. An application of ultramicrotitration and volume measurement methods to a quantitative study of the exopeptidases in skin sections." Arch Dermatol **82**: 595-604.

Hayes, T. L., N. Zimmerman, et al. (2006). Industry Market Research For Business Leaders, Strategists, Decision-Makers. Cleveland, OH, The Freedonia Group, Inc.

Helland, R., I. Leiros, et al. (1998). "The crystal structure of anionic salmon trypsin in complex with bovine pancreatic trypsin inhibitor." Eur J Biochem **256**(2): 317-24.

Hernandez-Cortes, P., L. Cerenius, et al. (1999). "Trypsin from Pacifastacus leniusculus hepatopancreas: purification and cDNA cloning of the synthesized zymogen." Biol Chem **380**(4): 499-501.

Hink-Schauer, C., E. Estebanez-Perpina, et al. (2003). "Crystal structure of the apoptosis-inducing human granzyme A dimer." Nat Struct Biol **10**(7): 535-40.

Hirsh, J. and R. Schleif (1977). "The araC promoter: transcription, mapping and interaction with the araBAD promoter." Cell **11**(3): 545-50.

Hjeltnes, B., K. Andersen, et al. (1987). "Experimental studies on the pathogenicity of a vibrio sp. isolated from Atlantic salmon, Salmo salar L., suffering from Hitra Disease." Journal of Fish Diseases **10**(1): 21-27.

Hjerde, E. (2007). the complete genome sequence of the fish pathogen *vibrio salmonicida.* Molecular biotechnology. tromsø, tromsø university. **Philoosphiae Doctor:** 28.

Hjerde, E., M. S. Lorentzen, et al. (2008). "The genome sequence of the fish pathogen Aliivibrio salmonicida strain LFI1238 shows extensive evidence of gene decay." BMC Genomics **9**: 616.

Hocman, G. (1992). "Chemoprevention of cancer: protease inhibitors." Int J Biochem **24**(9): 1365-75.

Hoff, K. A. (1989). "Survival of Vibrio anguillarum and Vibrio salmonicida at different salinities." Appl Environ Microbiol **55**(7): 1775-86.

Holm, K. and T. Jørgensen (1987). "A successful vaccination of Atlantic salmon, Salmo salar L., against 'Hitra disease' or coldwater vibriosis." Journal of Fish Diseases **10**(2): 85 - 90.

Holm, L. and J. Park (2000). "DaliLite workbench for protein structure comparison." Bioinformatics **16**(6): 566-7.

Hopp, T. P. and K. R. Woods (1981). "Prediction of protein antigenic determinants from amino acid sequences." Proc Natl Acad Sci U S A **78**(6): 3824-8.

Hrtley, B. and D. Kauffman (1996). "Corrections to the aminoacid sequence of Bovine Chymotrypsinsogen A. ." Biochem J **101**: 229-231.

Ikai, A. (1980). "Thermostability and aliphatic index of globular proteins." J Biochem **88**(6): 1895-8.

Irr, J. and E. Englesberg (1971). "Control of expression of the L-arabinose operon in temperature-sensitive mutants of gene araC in Escherichia coli B-r." J Bacteriol **105**(1): 136-41.

Ishida, M., T. Oshima, et al. (2002). "Overexpression in Escherichia coli of the AT-rich trpA and trpB genes from the hyperthermophilic archaeon Pyrococcus furiosus." FEMS Microbiol Lett **216**(2): 179-83.

Jacob, F. and J. Monod (1961). "Genetic regulatory mechanisms in the synthesis of proteins." J Mol Biol **3**: 318-56.

Jacob, F., D. Perrin, et al. (1960). "[Operon: a group of genes with the expression coordinated by an operator.]." C R Hebd Seances Acad Sci **250**: 1727-9.

Jana, S. and J. K. Deb (2005). "Strategies for efficient production of heterologous proteins in Escherichia coli." Appl Microbiol Biotechnol **67**(3): 289-98.

Kabsch, W. and C. Sander (1983). "How good are predictions of protein secondary structure?" FEBS Lett **155**(2): 179-82.

Kahn, M., R. Kolter, et al. (1979). "Plasmid cloning vehicles derived from plasmids ColE1, F, R6K, and RK2." Methods Enzymol **68**: 268-80.

Kane, J. F. (1995). "Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli." Curr Opin Biotechnol **6**(5): 494-500.

Kapust, R. B. and D. S. Waugh (1999). "Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused." Protein Sci **8**(8): 1668-74.

Kawabata, A. (2001). "[The G protein-coupled protease receptor PAR (protease-activated receptor) as a novel target for drug development]." Yakugaku Zasshi **121**(1): 1-7.

Kawasaki, H., Y. Kurosu, et al. (1986). "Limited digestion of calmodulin with trypsin in the presence or absence of various metal ions." J Biochem **99**(5): 1409-16.

Keevil, C. W., B. J. Spillane, et al. (1987). "Plasmid stability and antibiotic resistance of Neisseria gonorrhoea during glucose-limited continuous culture." J Med Microbiol **24**(4): 351-7.

Keil, B. (1992). Specificity of proteolysis. Berlin-Heidelberg-NewYork, Springer-Verlag.

Kelemen, B. R., T. A. Klink, et al. (1999). "Hypersensitive substrate for ribonucleases." Nucleic Acids Res **27**(18): 3696-701.

Kennedy, A. R. (1993). "Cancer prevention by protease inhibitors." Prev Med **22**(5): 796-811.

Kim, S., Y. Leem, et al. (2001). "Cloning of an acidic laccase gene (clac2) from Coprinus congregatus and its expression by external pH." FEMS Microbiol Lett **195**(2): 151-6.

Kim, S. W. and J. D. Keasling (2001). "Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in Escherichia coli enhances lycopene production." Biotechnol Bioeng **72**(4): 408-15.

Kishimura, H. and K. Hayashi (2002). "Isolation and characteristics of trypsin from pyloric ceca of the starfish Asterina pectinifera." Comp Biochem Physiol B Biochem Mol Biol **132**(2): 485-90.

Klein, B., G. Le Moullac, et al. (1996). "Molecular cloning and sequencing of trypsin cDNAs from Penaeus vannamei (Crustacea, Decapoda): use in assessing gene expression during the moult cycle." Int J Biochem Cell Biol **28**(5): 551-63.

Klose, J., N. Wendt, et al. (2004). "Hexa-histidin tag position influences disulfide structure but not binding behavior of in vitro folded N-terminal domain of rat corticotropin-releasing factor receptor type 2a." Protein Sci **13**(9): 2470-5.

Knight, C. G., P. M. Dando, et al. (1995). "Thimet oligopeptidase specificity: evidence of preferential cleavage near the C-terminus and product inhibition from kinetic analysis of peptide hydrolysis." Biochem J **308 (Pt 1)**: 145-50.

Kostakioti, M., C. L. Newman, et al. (2005). "Mechanisms of protein export across the bacterial outer membrane." J Bacteriol **187**(13): 4306-14.

Kumar, S. and R. Nussinov (2004). "Different roles of electrostatics in heat and in cold: adaptation by citrate synthase." Chembiochem **5**(3): 280-90.

Kyte, J. and R. F. Doolittle (1982). "A simple method for displaying the hydropathic character of a protein." J Mol Biol **157**(1): 105-32.

La Vallie, E. R., DiBlasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F. and McCoy, J. M. (1993).

Labaere, P., S. Gruenwald-Janho, et al. (2004). Roche Applied Sciences Lab FAQs. Purifying Proteins. U. Hoffmann-Rohrer and B. Kruchen, © 2003/2004 Roche Diagnostics Corporation. **2**.

Lala, P. K. and C. H. Graham (1990). "Mechanisms of trophoblast invasiveness and their control: the role of proteases and protease inhibitors." Cancer Metastasis Rev **9**(4): 369-79.

Lantz, M. S. (1997). "Are bacterial proteases important virulence factors?" J Periodontal Res **32**(1 Pt 2): 126-32.

Leaphart, A. B., D. K. Thompson, et al. (2006). "Transcriptome profiling of Shewanella oneidensis gene expression following exposure to acidic and alkaline pH." J Bacteriol **188**(4): 1633-42.

Lee, H. J., J. N. Larue, et al. (1971). "Angiotensin-converting enzyme from porcine plasma." Biochim Biophys Acta **235**(3): 521-8.

Lee, N., C. Francklyn, et al. (1987). "Arabinose-induced binding of AraC protein to araI2 activates the araBAD operon promoter." Proc Natl Acad Sci U S A **84**(24): 8814-8.

Lee, N., G. Wilcox, et al. (1974). "In vitro activation of the transcription of araBAD operon by araC activator." Proc Natl Acad Sci U S A **71**(3): 634-8.

Lee, W. S., C. H. Park, et al. (2004). "Streptomyces griseus trypsin is stabilized against autolysis by the cooperation of a salt bridge and cation-pi interaction." J Biochem **135**(1): 93-9.

Leiros, H. K., N. P. Willassen, et al. (1999). "Residue determinants and sequence analysis of cold-adapted trypsins." Extremophiles **3**(3): 205-19.

Levy, R., R. Weiss, et al. (2001). "Production of correctly folded Fab antibody fragment in the cytoplasm of Escherichia coli trxB gor mutants via the coexpression of molecular chaperones." Protein Expr Purif **23**(2): 338-47.

Li, M., G. S. Laco, et al. (2005). "Crystal structure of human T cell leukemia virus protease, a novel target for anticancer drug design." Proc Natl Acad Sci U S A **102**(51): 18332-7.

Lin-Chao, S., W. T. Chen, et al. (1992). "High copy number of the pUC plasmid results from a Rom/Rop-suppressible point mutation in RNA II." Mol Microbiol **6**(22): 3385-93.

Liu, Y., T. J. Zhao, et al. (2005). "Increase of soluble expression in Escherichia coli cytoplasm by a protein disulfide isomerase gene fusion system." Protein Expr Purif **44**(2): 155-61.

Lopes, A. R., M. A. Juliano, et al. (2006). "Substrate specificity of insect trypsins and the role of their subsites in catalysis." Insect Biochem Mol Biol **36**(2): 130-40.

Luan, C. H., S. Qiu, et al. (2004). "High-throughput expression of C. elegans proteins." Genome Res **14**(10B): 2102-10.

Luo, Z. H. and Z. C. Hua (1998). "Increased solubility of glutathione S-transferase-P16 (GST-p16) fusion protein by co-expression of chaperones groes and groel in Escherichia coli." Biochem Mol Biol Int **46**(3): 471-7.

Maeda, H. and A. Molla (1989). "Pathogenic potentials of bacterial proteases." Clin Chim Acta **185**(3): 357-67.

Makrides, S. C. (1996). "Strategies for achieving high-level expression of genes in Escherichia coli." Microbiol Rev **60**(3): 512-38.

Marsischky, G. and J. LaBaer (2004). "Many paths to many clones: a comparative look at high-throughput cloning methods." Genome Res **14**(10B): 2020-8.

Mary, A., K. E. Achyuthan, et al. (1988). "The binding of divalent metal ions to platelet factor XIII modulates its proteolysis by trypsin and thrombin." Arch Biochem Biophys **261**(1): 112-21.

Maxwell, K. L., A. K. Mittermaier, et al. (1999). "A simple in vivo assay for increased protein solubility." Protein Sci **8**(9): 1908-11.

Mayer, M. P. (1995). "A new set of useful cloning and expression vectors derived from pBlueScript." Gene **163**(1): 41-6.

McGowan, C. C., A. S. Necheva, et al. (2003). "Promoter analysis of Helicobacter pylori genes with enhanced expression at low pH." Mol Microbiol **48**(5): 1225-39.

McGrath, B. W., G (2005). Directory of Therapeutic Enzymes, CRC Press.

Mergulhao, F. J., G. A. Monteiro, et al. (2003). "Medium and copy number effects on the secretion of human proinsulin in Escherichia coli using the universal stress promoters uspA and uspB." Appl Microbiol Biotechnol **61**(5-6): 495-501.

Mergulhao, F. J., D. K. Summers, et al. (2005). "Recombinant protein secretion in Escherichia coli." Biotechnol Adv **23**(3): 177-202.

Miot, M. and J. M. Betton (2004). "Protein quality control in the bacterial periplasm." Microb Cell Fact **3**(1): 4.

Miyagawa, S., N. Nishino, et al. (1991). "Effects of protease inhibitors on growth of Serratia marcescens and Pseudomonas aeruginosa." Microb Pathog **11**(2): 137-41.

Mohanty, A. K. and M. C. Wiener (2004). "Membrane protein expression and production: effects of polyhistidine tag length and position." Protein Expr Purif **33**(2): 311-25.

Morris, A. L., M. W. MacArthur, et al. (1992). "Stereochemical quality of protein structure coordinates." Proteins **12**(4): 345-64.

Muhlia-Almazan, A., A. Sanchez-Paz, et al. (2008). "Invertebrate trypsins: a review." J Comp Physiol [B] **178**(6): 655-72.

Muller, H. M., J. M. Crampton, et al. (1993). "Members of a trypsin gene family in Anopheles gambiae are induced in the gut by blood meal." Embo J **12**(7): 2891-900.

Nagai, K., H. C. Thogersen, et al. (1988). "Refolding and crystallographic studies of eukaryotic proteins produced in Escherichia coli." Biochem Soc Trans **16**(2): 108-10.

Nallamsetty, S., B. P. Austin, et al. (2005). "Gateway vectors for the production of combinatorially-tagged His6-MBP fusion proteins in the cytoplasm and periplasm of Escherichia coli." Protein Sci **14**(12): 2964-71.

Nallamsetty, S. and D. S. Waugh (2006). "Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners." Protein Expr Purif **45**(1): 175-82.

Nardi, G. L. (1960). "Serum "trypsin" (or arginine exopeptidase) screening test for cancer of the panceas." Gastroenterology **38**: 50-1.

Nicholas, K. and H. Nicholas (1997). GeneDoc: Atool for editing and annotating multiple sequence alignment.

Niiranen, L., B. Altermark, et al. (2008). "Effects of salt on the kinetics and thermodynamic stability of endonuclease I from Vibrio salmonicida and Vibrio cholerae." Febs J **275**(7): 1593-605.

Niiranen, L., S. Espelid, et al. (2007). "Comparative expression study to increase the solubility of cold adapted Vibrio proteins in Escherichia coli." Protein Expr Purif **52**(1): 210-8.

Nilsson, T., J. Carlsson, et al. (1985). "Inactivation of key factors of the plasma proteinase cascade systems by Bacteroides gingivalis." Infect Immun **50**(2): 467-71.

Nomine, Y., T. Ristriani, et al. (2001). "A strategy for optimizing the monodispersity of fusion proteins: application to purification of recombinant HPV E6 oncoprotein." Protein Eng **14**(4): 297-305.

Novagen (2003). pET system manual.

Nuc, P. and K. Nuc (2006). "[Recombinant protein production in Escherichia coli]." Postepy Biochem **52**(4): 448-56.

O'Halloran, J. (1993). "Additional information about the occurrence of Hitra disease in Atlantic salmon." Can Vet J **34**(1): 6.

Olins, P. O. and S. H. Rangwala (1989). "A novel sequence element derived from bacteriophage T7 mRNA acts as an enhancer of translation of the lacZ gene in Escherichia coli." J Biol Chem **264**(29): 16973-6.

Osawa, S., T. Ohama, et al. (1989). "Evolution of the mitochondrial genetic code. I. Origin of AGR serine and stop codons in metazoan mitochondria." J Mol Evol **29**(3): 202-7.

Osawa, S., T. Ohama, et al. (1989). "Evolution of the mitochondrial genetic code. II. Reassignment of codon AUA from isoleucine to methionine." J Mol Evol **29**(5): 373-80.

Ossovskaya, V. S. and N. W. Bunnett (2004). "Protease-activated receptors: contribution to physiology and disease." Physiol Rev **84**(2): 579-621.

Parsons, M. E. and R. J. Pennington (1976). "Separation of rat muscle aminopeptidases." Biochem J **155**(2): 375-81.

Pawelczyk, E., M. Zajac, et al. (1981). "Kinetics of drug decomposition. Part 66. Kinetics of the hydrolysis of carphecillin in aqueous solution." Pol J Pharmacol Pharm **33**(3): 373-86.

Perona, J. J. and C. S. Craik (1995). "Structural basis of substrate specificity in the serine proteases." Protein Sci **4**(3): 337-60.

Perryman, A. L., J. H. Lin, et al. (2006). "Restrained molecular dynamics simulations of HIV-1 protease: the first step in validating a new target for drug design." Biopolymers **82**(3): 272-84.

Puente, X. S., L. M. Sanchez, et al. (2003). "Human and mouse proteases: a comparative genomic approach." Nat Rev Genet **4**(7): 544-58.

Paal, M., T. Heel, et al. (2009). "A novel Ecotin-Ubiquitin-Tag (ECUT) for efficient, soluble peptide production in the periplasm of Escherichia coli." Microb Cell Fact **8**: 7.

Raines, R. T., M. McCormick, et al. (2000). "The S.Tag fusion system for protein purification." Methods Enzymol **326**: 362-76.

Rawlings, N. D. and A. J. Barrett (1993). "Evolutionary families of peptidases." Biochem J **290 (Pt 1)**: 205-18.

Rawlings, N. D. and A. J. Barrett (1994). "Families of cysteine peptidases." Methods Enzymol **244**: 461-86.

Rawlings, N. D. and A. J. Barrett (1995). "Evolutionary families of metallopeptidases." Methods Enzymol **248**: 183-228.

Rawlings, N. D. and A. J. Barrett (1995). "Families of aspartic peptidases, and those of unknown catalytic mechanism." Methods Enzymol **248**: 105-20.

Rawlings, N. D. and A. J. Barrett (1997). "Structure of membrane glutamate carboxypeptidase." Biochim Biophys Acta **1339**(2): 247-52.

Rawlings, N. D. and A. J. Barrett (1999). "MEROPS: the peptidase database." Nucleic Acids Res **27**(1): 325-31.

Rawlings, N. D. and A. J. Barrett (2000). "MEROPS: the peptidase database." Nucleic Acids Res **28**(1): 323-5.

Rawlings, N. D., F. R. Morton, et al. (2006). "MEROPS: the peptidase database." Nucleic Acids Res **34**(Database issue): D270-2.

Rawlings, N. D., E. O'Brien, et al. (2002). "MEROPS: the protease database." Nucleic Acids Res **30**(1): 343-6.

Rawlings, N. D., D. P. Tolle, et al. (2004). "MEROPS: the peptidase database." Nucleic Acids Res **32**(Database issue): D160-4.

Reeck, G. R., C. de Haen, et al. (1987). ""Homology" in proteins and nucleic acids: a terminology muddle and a way out of it." Cell **50**(5): 667.

Rennard, S. I., K. Rickard, et al. (1991). "Protease injury in airways disease." Ann N Y Acad Sci **624**: 278-85.

Rochefort, H., F. Capony, et al. (1990). "Cathepsin D: a protease involved in breast cancer metastasis." Cancer Metastasis Rev **9**(4): 321-31.

Rochefort, H. and E. Liaudet-Coopman (1999). "Cathepsin D in cancer metastasis: a protease and a ligand." Apmis **107**(1): 86-95.

Rowan, A. D., D. J. Buttle, et al. (1990). "The cysteine proteinases of the pineapple plant." Biochem J **266**(3): 869-75.

Rypniewski, W. R., P. R. Ostergaard, et al. (2001). "Fusarium oxysporum trypsin at atomic resolution at 100 and 283 K: a study of ligand binding." Acta Crystallogr D Biol Crystallogr **57**(Pt 1): 8-19.

Sachdev, D. and J. M. Chirgwin (1998). "Order of fusions between bacterial and mammalian proteins can determine solubility in Escherichia coli." Biochem Biophys Res Commun **244**(3): 933-7.

Sachdev, D. and J. M. Chirgwin (1999). "Properties of soluble fusions between mammalian aspartic proteinases and bacterial maltose-binding protein." J Protein Chem **18**(1): 127-36.

Saejung, W., C. Puttikhunt, et al. (2006). "Enhancement of recombinant soluble dengue virus 2 envelope domain III protein production in Escherichia coli trxB and gor double mutant." J Biosci Bioeng **102**(4): 333-9.

Saier, M. H., Jr. (1995). "Differential codon usage: a safeguard against inappropriate expression of specialized genes?" FEBS Lett **362**(1): 1-4.

Saji, T. (2008). "[Clinical utility of ulinastatin, urinary protease inhibitor in acute Kawasaki disease]." Nippon Rinsho **66**(2): 343-8.

Salte, R., P. Nafstad, et al. (1987). "Disseminated intravascular coagulation in "Hitra disease" (hemorrhagic syndrome) in farmed Atlantic salmon." Vet Pathol **24**(5): 378-85.

Sambrook, J., E. F. Fritsch, et al. (1989). Molecular Cloning. A Laboratory Manual. NY, Cold Spring Harbor Laboratory Press

San, K. Y., G. N. Bennett, et al. (1994). "An optimization study of a pH-inducible promoter system for high-level recombinant protein production in Escherichia coli." Ann N Y Acad Sci **721**: 268-76.

Schechter, I. and A. Berger (1967). "On the size of the active site in proteases. I. Papain." Biochem Biophys Res Commun **27**(2): 157-62.

Schmidt, T., K. Friehs, et al. (1996). "Rapid determination of plasmid copy number." J Biotechnol **49**(1-3): 219-29.

Schwarz, D., V. Dotsch, et al. (2008). "Production of membrane proteins using cell-free expression systems." Proteomics **8**(19): 3933-46.

Seif, C., S. Boy, et al. (2008). "[Botulinum toxin for the treatment of overactive bladder - an overview.]." Urologe A **47**(1): 46-53.

Seife, C. (1997). "Blunting nature's Swiss army knife." Science **277**(5332): 1602-3.

Sharp, P. M., M. Stenico, et al. (1993). "Codon usage: mutational bias, translational selection, or both?" Biochem Soc Trans **21**(4): 835-41.

Shokri, A., A. M. Sanden, et al. (2003). "Cell and process design for targeting of recombinant protein into the culture medium of Escherichia coli." Appl Microbiol Biotechnol **60**(6): 654-64.

Siddiqui, K. S. and R. Cavicchioli (2006). "Cold-adapted enzymes." Annu Rev Biochem **75**: 403-33.

Sievert, K. D., J. Bremer, et al. (2007). "[Botulinum toxin for the treatment of neurogenic detrusor hyperactivity. Consensus paper on use for neurogenic bladder dysfunction]." Urologe A **46**(3): 293-6.

Simonen, M. and I. Palva (1993). "Protein secretion in Bacillus species." Microbiol Rev **57**(1): 109-37.

Simonovic, I. and P. A. Patston (2000). "The native metastable fold of C1-inhibitor is stabilized by disulfide bonds." Biochim Biophys Acta **1481**(1): 97-102.

Simons, R. W., F. Houman, et al. (1987). "Improved single and multicopy lac-based cloning vectors for protein and operon fusions." Gene **53**(1): 85-96.

Slocombe, B. and R. Sutherland (1969). "Beta-lactamase activity and resistance to ampicillin, carbenicillin, and cephaloridine of Klebsiella, Enterobacter, and Citrobacter." Antimicrob Agents Chemother (Bethesda) **9**: 78-85.

Southan, C. (2000). "Assessing the protease and protease inhibitor content of the human genome." J Pept Sci **6**(9): 453-8.

Staunton, D., R. Schlinkert, et al. (2006). "Cell-free expression and selective isotope labelling in protein NMR." Magn Reson Chem **44 Spec No**: S2-9.

Stephen, C. (1993). "Questions about Hitra disease in Atlantic salmon." Can Vet J **34**(1): 5-6.

Stoker, N. G., N. F. Fairweather, et al. (1982). "Versatile low-copy-number plasmid vectors for cloning in Escherichia coli." Gene **18**(3): 335-41.

Studier, F. W. (1991). "Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system." J Mol Biol **219**(1): 37-44.

Studier, F. W. (2005). "Protein production by auto-induction in high density shaking cultures." Protein Expr Purif **41**(1): 207-34.

Sugimoto, S., T. Fujii, et al. (2007). "The fibrinolytic activity of a novel protease derived from a tempeh producing fungus, Fusarium sp. BLB." Biosci Biotechnol Biochem **71**(9): 2184-9.

Tachibana, A., K. Tohiguchi, et al. (2009). "Preparation of long sticky ends for universal ligation-independent cloning: Sequential T4 DNA polymerase treatments." J Biosci Bioeng **107**(6): 668-9.

Tanford, C. (1958). Physical Chemistry of macro molecules. New York, Jhon Wiley & Sons.

Taylor, P., V. Anderson, et al. (1999). "Novel mechanism of inhibition of elastase by beta-lactams is defined by two inhibitor crystal complexes." J Biol Chem **274**(35): 24901-5.

Thompson, J. D., T. J. Gibson, et al. (2002). "Multiple sequence alignment using ClustalW and ClustalX." Curr Protoc Bioinformatics **Chapter 2**: Unit 2 3.

Thompson, J. D., T. J. Gibson, et al. (1997). "The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools." Nucleic Acids Res **25**(24): 4876-82.

Thompson, J. D., D. G. Higgins, et al. (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Res **22**(22): 4673-80.

Thorvaldsen, S., E. Hjerde, et al. (2007). "Molecular characterization of cold adaptation based on ortholog protein sequences from Vibrionaceae species." Extremophiles **11**(5): 719-32.

Timmer, J. C. and G. S. Salvesen (2007). "Caspase substrates." Cell Death Differ **14**(1): 66-72.

Tipton, K. F. (1994). "Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). Enzyme nomenclature. Recommendations 1992. Supplement: corrections and additions." Eur J Biochem **223**(1): 1-5.

Titani, K., T. Sasagawa, et al. (1983). "Amino acid sequence of crayfish (Astacus fluviatilis) trypsin If." Biochemistry **22**(6): 1459-65.

Tobias, J. W., T. E. Shrader, et al. (1991). "The N-end rule in bacteria." Science **254**(5036): 1374-7.

Travis, J., J. Potempa, et al. (1995). "Are bacterial proteinases pathogenic factors?" Trends Microbiol **3**(10): 405-7.

Tripathi, L. P. and R. Sowdhamini (2008). "Genome-wide survey of prokaryotic serine proteases: analysis of distribution and domain architectures of five serine protease families in prokaryotes." BMC Genomics **9**: 549.

Turner, P., O. Holst, et al. (2005). "Optimized expression of soluble cyclomaltodextrinase of thermophilic origin in Escherichia coli by using a soluble fusion-tag and by tuning of inducer concentration." Protein Expr Purif **39**(1): 54-60.

Uhlin, B. E., V. Schweickart, et al. (1983). "New runaway-replication-plasmid cloning vectors and suppression of runaway replication by novobiocin." Gene **22**(2-3): 255-65.

Vermersch, P. S., M. R. Klass, et al. (1986). "Use of bacterial DHFR-II fusion proteins to elicit specific antibodies." Gene **41**(2-3): 289-97.

Villalonga, M. L., G. Reyes, et al. (2004). "Metal-induced stabilization of trypsin modified with alpha-oxoglutaric acid." Biotechnol Lett **26**(3): 209-12.

Waldo, G. S., B. M. Standish, et al. (1999). "Rapid protein-folding assay using green fluorescent protein." Nat Biotechnol **17**(7): 691-5.

Walhout, A. J., G. F. Temple, et al. (2000). "GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes." Methods Enzymol **328**: 575-92.

Walsh, K. A. (1970). "Trypsinogens and Trypsins of various species." Methods Enzymol **19**: 41-46.

Wefer, B., C. Seif, et al. (2007). "[Botulinum toxin A injection for treatment-refractory giggle incontinence]." Urologe A **46**(7): 773-5.

Wigley, W. C., R. D. Stidham, et al. (2001). "Protein solubility and folding monitored in vivo by structural complementation of a genetic marker protein." Nat Biotechnol **19**(2): 131-6.

Wild, J. and W. Szybalski (2004). "Copy-control pBAC/oriV vectors for genomic cloning." Methods Mol Biol **267**: 145-54.

Wild, J. and W. Szybalski (2004). "Copy-control tightly regulated expression vectors based on pBAC/oriV." Methods Mol Biol **267**: 155-67.

Wilkins, J. C., D. Beighton, et al. (2003). "Effect of acidic pH on expression of surface-associated proteins of Streptococcus oralis." Appl Environ Microbiol **69**(9): 5290-6.

Wilkinson, D. L. and R. G. Harrison (1991). "Predicting the solubility of recombinant proteins in Escherichia coli." Biotechnology (N Y) **9**(5): 443-8.

Wilmouth, R. C., S. Kassamally, et al. (1999). "Mechanistic insights into the inhibition of serine proteases by monocyclic lactams." Biochemistry **38**(25): 7989-98.

Wilms, B., A. Hauck, et al. (2001). "High-cell-density fermentation for production of L-N-carbamoylase using an expression system based on the Escherichia coli rhaBAD promoter." Biotechnol Bioeng **73**(2): 95-103.

Withers-Martinez, C., E. P. Carpenter, et al. (1999). "PCR-based gene synthesis as an efficient approach for expression of the A+T-rich malaria genome." Protein Eng **12**(12): 1113-20.

Wolstenholme, D. R. (1992). "Animal mitochondrial DNA: structure and evolution." Int Rev Cytol **141**: 173-216.

Wong, J. H., N. Cai, et al. (2004). "Thioredoxin reduction alters the solubility of proteins of wheat starchy endosperm: an early event in cereal germination." Plant Cell Physiol **45**(4): 407-15.

Wu, Q. (2007). "The serine protease corin in cardiovascular biology and disease." Front Biosci **12**: 4179-90.

Yang, J., X. Huang, et al. (2005). "Isolation and characterization of a serine protease from the nematophagous fungus, Lecanicillium psalliotae, displaying nematicidal activity." Biotechnol Lett **27**(15): 1123-8.

Yasukawa, T., C. Kanei-Ishii, et al. (1995). "Increase of solubility of foreign proteins in Escherichia coli by coproduction of the bacterial thioredoxin." J Biol Chem **270**(43): 25328-31.

Young, C., K. Matsubara, et al. (1996). "A modified unique site elimination mutagenesis in constructing a chloramphenicol resistance-encoding pGEM vector." Biotechniques **20**(6): 986-8.

Yu, P., A. Aristidou, et al. (1991). "Synergistic effect of glycine and bacteriocin release protein in the release of periplasmic protein in recombinantE. coli " Journal Biotechnology Letters **13**(5): 311-316.

Yu, P., A. A. Aristidou, et al. (1991). "synergetic effects of glycine and bacteriocin relase protein in the recombinant *E. coli.*" biotechnology letters **13**(5): 311-16.

Zhang, Y., L. Taiming, et al. (2003). "Low temperature and glucose enhanced T7 RNA polymerase-based plasmid stability for increasing expression of glucagon-like peptide-2 in Escherichia coli." Protein Expr Purif **29**(1): 132-9.

Zhang, Z., Z. H. Li, et al. (2002). "Overexpression of DsbC and DsbG markedly improves soluble and functional expression of single-chain Fv antibodies in Escherichia coli." Protein Expr Purif **26**(2): 218-28.

Appendices

Table 7.1: Bioinformatic analysis of Aliivibrio salmonicida genome

Vs ORFp	Length of ORF	Blast with MEROPS		(Blast with NCBI nr)							CD results		Blast with PDB	
					Highest similarity found with									
		Peptide Family	Ref	Other info	Protein	Organism	% Id	%+Ve	Score	E-value	code	% aligned	PDB code	%Positive
1.	174	A8	37VSmp	gi|59479176|gb|AAW84963.1|	lipoprotein signal peptidase	Vibrio fischeri ES114	159/171 (92%)	162/171 (94%)	244	8e-64	-------------	---------------	gi|48425542|pdb|1S2W|A	24/46 (52%),
2.	259	M16B	47VSmp	gi|50748728|ref|XP_421380.1|	similar to KIAA1124 protein	Gallus gallus	52/220 (23%),	103/220 (46%),	43.9	0.006	gnl|CDD|14075 Membrane-bound metallopeptidase	39.0%	gi|62738836|pdb|1YYC|A	21/44 (47%),
3.	449	S16	50VSmp	gi|59711900|ref|YP_204676.1| **LonB**	ATP-dependent protease La	Vibrio fischeri ES114	324/437 (74%),	375/437 (85%),	584	2e-165	gnl|CDD|10791	57.2%	gi|55670858|pdb|1XHK|B	64/133 (48%),
4.	158	S16	50VSmp	gi|59711900|ref|YP_204676.1| **Lon_C**	ATP-dependent protease La	Vibrio fischeri ES114	92/118 (77%),	107/118 (90%),	206	2e-52	gnl|CDD|23606	29.3%	gi|73535793|pdb|1Z0W|A	36/69 (52%),
5.	515	S13	55VSmp	gi|59711081|ref|YP_203857.1|	D-alanyl-meso-diaminopimelate endopeptidase	Vibrio fischeri ES114	331/448 (73%),	391/448 (87%),	637	0.0	gnl|CDD|11735	93.4%	gi|71041813|pdb|1W8Y|D	192/451 (42%),
6.	744	M44	58VSmp	(1).gi|66048002|ref|YP_237843.1|	hypothetical protein Psyr_4778	Pseudomonas syringae pv. syringae B728a	37/119 (31%),	54/119 (45%),	64.7	1e-08	-----------	---------------	gi|66361586|pdb|2BO9|D --------- gi|61680281|pdb|1WXR|A (heam protease)	39/100 (39%), 19/40 (47%),
7.	244	M44	58VSmp	gi|71363631|ref|ZP_00654235.1|	Collagen triple helix repeat:Haemagluttinin motif:Hep_Hag	Psychrobacter cryohalolentis K5	43/152 (28%),	68/152 (44%),	37.0	0.78	gnl|CDD|8822	100.0%	gi|83754122|pdb|2AWB|E	30/58 (51%),
8.	361	M50B	62VSmp	gi|59711816|ref|YP_204592.1|	membrane metalloprotease	Vibrio fischeri ES114	317/360 (88%),	342/360 (95%),	510	4e-143	gnl|CDD|11702	60.9%	---------------------	----------------------------

9.	357	S49	71VSmp	gi|59711647|ref|YP_204423.1|	possible protease SohB	Vibrio fischeri ES114	324/353 (91%),	340/353 (96%),	484	3e-135	?????	????	gi|73535949|pdb|1ZKG|B	22/35 (62%),
10.	711	M41	72VSmp	gi|59711086|ref|YP_203862.1| Peptidase family M41	cell division protein FtsH	Vibrio fischeri ES114	584/611 (95%),	600/611 (98%),	1090	0.0	gnl|CDD|10338	100.0%	gi|90109150|pdb|2CEA|F	339/451 (75%),
11.	513	A8	74VSmp	gi|59713256|ref|YP_206031.1|	di-/tripeptide transporter	Vibrio fischeri ES114	443/482 (91%),	464/482 (96%),	751	0.0	gnl|CDD|12443	92.4%	gi|34811518|pdb|1PF4|D	31/58 (53%),
12.	802	S1C	74VSmp	gi|75829692|ref|ZP_00758987.1|	Large exoproteins involved in heme utilization or adhesion	Vibrio cholerae MO10	310/820 (37%),	426/820 (51%),	431	7e-119	----------- -----------	‘’’’’’’’’’’’’’’ ’’’’’’’’’’’’’’’ ’	gi|1942464|pdb|1IGN|B	21/35 (60%),
13.	284	S54	78VSmp	gi|59713054|ref|YP_205830.1|	integral membrane protein (rhomboid family)	Vibrio fischeri ES114	241/275 (87%),	260/275 (94%),	453	4e-126	gnl|CDD|10574	88.6%	gi|20664086|pdb|1KEN|E	31/72 (43%),
14.	208	S24	78VSmp	gi|59713049|ref|YP_205825.1| Peptidase_S24	LexA repressor	Vibrio fischeri ES114	190/208 (91%),	199/208 (95%),	322	5e-87	gnl|CDD|4309	98.2%	gi|15988321|pdb|1JHH|B	173/207 (83%),
15.	495	M20C	79VSmp	gi|59711343|ref|YP_204119.1|	aminoacyl-histidine dipeptidase	Vibrio fischeri ES114	454/484 (93%),	474/484 (97%),	944	0.0	gnl|CDD|11903	100.0%	gi|40890020|pdb|1VIX|B	35/59 (59%),
16.	459	S9C	79VSmp	gi|71366653|ref|ZP_00657192.1|	Dipeptidyl aminopeptidases /acylaminoacyl-peptidases-like	Acidothermus cellulolyticus 11B	54/207 (26%),	90/207 (43%),	66.2	3e-09	gnl|CDD|11220	26.6%	gi|56554274|pdb|1VE7|B	55/136 (40%),
17.	337	U32	85VSmp	gi|59711100|ref|YP_203876.1|	putative protease YhbU precursor	Vibrio fischeri ES114	322/333 (96%),	329/333 (98%),	676	0.0	gnl|CDD|23156	100.0%	gi|38492421|pdb|1KFK|A	48/106 (45%),
18.	295	U32	85VSmp	gi|59711099|ref|YP_203875.1|	putative protease YhbV precursor	Vibrio fischeri ES114	264/292 (90%),	277/292 (94%),	560	3e-158	gnl|CDD|23156	91.2%	gi|28373302|pdb|1GWI|B	43/103 (41%),
19.	622	S49	86VSmp	gi|59712250|ref|YP_205026.1|	signal peptide peptidase SppA	Vibrio fischeri ES114	548/618 (88%),	587/618 (94%),	1081	0.0	gnl|CDD|10486	97.8% 83.6%	gi|62738702|pdb|1YOV|C	53/122 (43%),
20.	483	S6	87VSmp	gi|77961243|ref|ZP_00825086.1|	COG3210: Large exoproteins involved in heme utilization or adhesion	Yersinia mollaretii ATCC 43969	100/404 (24%),	165/404 (40%),	88.2	7e-16	----------- ------	----------- ----------- ---	gi|14488780|pdb|1J71|A	67/154 (43%),

21.	349	M23B	88VSmp	gi|59711162|ref|YP_203938.1| Peptidase family M23/M37	(ToxR-activated protein) TagE-like gene	Vibrio fischeri ES114	290/337 (86%),	320/337 (94%),	583	4e-165	gnl|CDD|2090	100.0%	gi|88192196|pdb|2B13|B	46/86 (53%),
22.	581	M9B	90VSmp	gi|37677332|ref|NP_937728.1|	putative exoprotease Vcc	Vibrio vulnificus YJ016	285/580 (49%),	405/580 (69%),	562	2e-158	gnl|CDD|23266	100.0%	gi|18655953|pdb|1JLR|C	19/27 (70%),
23.	250	M9A	90VSmp	gi|3142333|gb|AAC23708.1|	metalloprotease	Vibrio mimicus	84/237 (35%),	129/237 (54%),	143	7e-33	--------------	-------------------	gi|83754308|pdb|2B8Q|F	26/53 (49%),
24.	686	S15	91VSmp	gi|13474275|ref|NP_105843.1| X-Pro dipeptidyl-peptidase	similar to glutaryl 7-ACA acylase	Mesorhizobium loti MAFF303099	308/654 (47%),	412/654 (62%),	616	1e-174	gnl|CDD|25875	94.3%	gi|18158643|pdb|1JU4|A	221/530 (41%),
25.	144	A24B	100VSmp	gi|86147182|ref|ZP_01065498.1| Flp pilus assembly protein, protease CpaA	hypothetical protein MED222_18936	Vibrio sp. MED222	60/141 (42%),	84/141 (59%),	74.7	1e-12	gnl|CDD|14092	51.8%	________	______________
26.	260	U32	102VSmp	gi|46913377|emb|CAG20165.1|	putative collagenase family protease	Photobacterium profundum	211/233 (90%),	224/233 (96%),	374	2e-102	gnl|CDD|23156	69.6%	gi|39655000|pdb|1US5|A	34/77 (44%),
27.	1000	M16B	105VSmp	gi|59711149|ref|YP_203925.1|	peptidase family M16	Vibrio fischeri ES114	826/950 (86%),	906/950 (95%),	1703	0.0	gnl|CDD|10753 Secreted/periplasmic Zn-dependent peptidases, insulinase-like	92.0%	gi|15826326|pdb|1HR6|H	184/462 (39%),
28.	71	S1A	106VSmp	gi|1906318|emb|CAA65252.1|	trypsinogen	Botryllus schlosseri	19/51 (37%),	29/51 (56%),	32.7	3.8	-----------------	-----------------	----------------	----------------
29.	385	M23B	108VSmp	gi|59712958|ref|YP_205734.1|	cell wall endopeptidase, family M23/M37	Vibrio fischeri ES114	324/381 (85%),	351/381 (92%),	541	2e-152	gnl|CDD|14075	95.7%	gi|88192196|pdb|2B13|B	55/124 (44%),
30.	132	M15C	114VSmp	gi|59712624|ref|YP_205400.1|	L-alanyl-D-glutamate peptidase	Vibrio fischeri ES114	98/128 (76%),	109/128 (85%),	187	1e-46	-----------------	-----------------	gi|88191788|pdb|1XP2|C	38/101 (37%),
31.	444	S11	117VSmp	gi|59711352|ref|YP_204128.1|	D-alanyl-D-alanine serine-type carboxypeptidase	Vibrio fischeri ES114	374/391 (95%),	387/391 (98%),	717	0.0	gnl|CDD|11397	96.2%	gi|71042191|pdb|1Z6F|A	253/346 (73%),
32.	414	M20B	118VSmp	gi|59713436|ref|YP_206211.1|	peptidase T	Vibrio fischeri ES114	377/409 (92%),	396/409 (96%),	745	0.0	gnl|CDD|11903	99.0%	gi|40890020|pdb|1VIX|B	316/411 (76%),
33.	363	C56	122VSm	gi|28900924|ref|NP_8	putative	Vibrio	73/191	106/191	114	9e-24	gnl|CDD|	100.0%	gi|56966619|	88/179

			p	00579.1| Tetratricopeptide repeat domain	virulence-mediating protein VirC	parahaemolyticus RIMD 2210633	(38%),	(55%),		1e-23	30164		pdb|1W25|B	(49%),
34.	1562	S6	122VSmp	gi|86144279|ref|ZP_01062611.1| Large exoproteins involved in heme utilization or adhesion	probable RTX (repeat in structural toxin)	Vibrio sp. MED222	679/1146 (59%),	822/1146 (71%),	993	0.0	-----------------	-----------------	gi|3745820|pdb|1SXZ|B	21/34 (61%),
35.	684	M3A	127VSmp	gi|59481195|gb|AAW86982.1|	Zn-dependent oligopeptidase A	Vibrio fischeri ES114	627/680 (92%),	655/680 (96%),	1233	0.0	gnl|CDD|10213	100.0%	gi|67463890|pdb|1Y79|1	346/685 (50%),
36.	507	M17	128VSmp	gi|59711020|ref|YP_203796.1|	leucyl aminopeptidase	Vibrio fischeri ES114	477/502 (95%),	494/502 (98%),	956	0.0	gnl|CDD|29554	100.0%	gi|21730302|pdb|1GYT|L	454/498 (91%),
37.	2401	M24A	130VSmp	gi|90406967|ref|ZP_01215158.1| blood coagulation protein von Willebrand factor (vWF)"	putative RTX toxin	Psychromonas sp. CNPT3	203/556 (36%),	306/556 (55%),	233	8e-59	gnl|CDD|16158	41.8%	gi|83754243|pdb|2B63|A	42/96 (43%),
38.	1245	S33	130VSmp	gi|90410085|ref|ZP_01218102.1| N-terminal double-glycine peptidase domain	hypothetical protein P3TCK_04941	Photobacterium profundum 3TCK	393/710 (55%),	543/710 (76%),	711	0.0	gnl|CDD|13766	89.5%	gi|34811518|pdb|1PF4|D	127/282 (45%),
39.	283	S9C	132VSmp	gi|69953478|ref|ZP_00640589.1|	putative esterase	Shewanella frigidimarina NCIMB 400	189/280 (67%),	229/280 (81%),	375	9e-103	gnl|CDD|10497	98.1%	gi|56965884|pdb|1PV1|D	180/293 (61%),
40.	859	S1D	132VSmp	gi|90021349|ref|YP_527176.1|	Peptidase M, neutral zinc metallopeptidases, zinc-binding site	Saccharophagus degradans 2-40	223/318 (70%),	262/318 (82%),	451	7e-125	gnl|CDD|12797	97.0%	gi|1942537|pdb|2HVM|	97/251 (38%),
41.	378	M20A	133VSmp	gi|59712914|ref|YP_205690.1| Peptidase family M20/M25/M40	acetylornithine deacetylase	Vibrio fischeri ES114	355/378 (93%),	371/378 (98%),	699	0.0	gnl|CDD|10494	92.7%	gi|39655034|pdb|1VGY|B	130/309 (42%),
42.	228	A2D	135VSmp	gi|59481715|gb|AAW87354.1|	ATP-dependent Zn proteases	Vibrio fischeri ES114	178/230 (77%),	189/230 (82%),	337	4e-91	gnl|CDD|23842	99.3%	gi|88192379|pdb|2B9P|E	16/24 (66%),
43.	461	S9C	140VSmp	gi|59711569|ref|YP_204345.1| Peptidase_S9_N	TolB protein precursor	Vibrio fischeri ES114	411/450 (91%),	437/450 (97%),	842	0.0	gnl|CDD|10690	99.1%	gi|12084602|pdb|1C5K|A	307/435 (70%),
44.	288	U48	140VSmp	gi|59713314|ref|YP_206089.1|	CAAX amino terminal protease family	Vibrio fischeri ES114	224/275 (81%),	250/275 (90%),	313	4e-84	gnl|CDD|8329	72.3%	----------------------	----------------

45.	414	M20B	141VSmp	gi|59479768|gb|AAW85555.1|	tripeptidase T	Vibrio fischeri ES114	369/409 (90%),	390/409 (95%),	734	0.0	gnl|CDD|11903	99.8%	gi|40890020|pdb|1VIX|B	235/402 (58%),
46.	707	S1X	145VSmp	gi|1073752|pir||B28534peptidase M4 and M36	luxC 5'-region	Vibrio harveyi	89/132 (67%)	106/132 (80%),	49.7	5e-04	XXXXXXXXXXXXXXX	XXXXXXXXXXXXXXX	gi|48425191|pdb|1P3E|A	42/95 (44%),
47.	301	A24A	147VSmp	gi|59712795|ref|YP_205571.1|	type 4 prepilin peptidase	Vibrio fischeri ES114	257/295 (87%),	274/295 (92%),	413	6e-114	gnl|CDD|11697	98.0%	gi|83754867|pdb|2CU8|A	20/38 (52%),
48.	213	S54	148VSmp	gi|69949016|ref|ZP_00637076.1| novel intramembrane serine protease	Rhomboid-like protein	Shewanella frigidimarina NCIMB 400	71/165 (43%),	102/165 (61%),	115	1e-24	gnl|CDD|25804	97.3%	gi|10120524|pdb|1F0C|A	23/39 (58%),
49.	512	S16	149VSmp	gi|84393426|ref|ZP_00992183.1| AAA-superfamily of ATPases associated with proteolysis etc	ComM-related protein	Vibrio splendidus 12B01	408/508 (80%),	443/508 (87%),	795	0.0	gnl|CDD|10476	100.0%	gi|15825870|pdb|1G8P|A	98/213 (46%),
50.	460	S16	151VSmp	gi|59479236|gb|AAW85023.1| (Lon protease (S16) C-terminal proteoly-tic domain)	DNA repair protein RadA	Vibrio fischeri ES114	441/460 (95%)	453/460 (98%)	868	0.0	gnl|CDD|10790	100.0%	gi|40889711|pdb|1RR9|F	49/89 (55%),
51.	3362	S6	152VSmp	gi|90407929|ref|ZP_01216103.1|	hemolysin-related protein	Psychromonas sp. CNPT3	1701/3363 (50%)	2274/3363 (67%)	2737	0.0	XXXXXXXXXXXXXXX	XXXXXXXXXXXXXXX	gi|83755027|pdb|2F17|B	22/44 (50%),
52.	203	C56	153VSmp	gi|75856136|ref|ZP_00763770.1| glutamine amidotransferase -GATase1	Putative intracellular protease/amidase	Vibrio sp. Ex25	120/195 (61%)	157/195 (80%)	249	4e-65	gnl|CDD|10562	100.0%	gi|42543006|pdb|1J42|A	107/189 (56%),
53.	382	M20A	155VSmp	gi|59480622|gb|AAW86409.1| Peptidase family M20/M25/M40	succinyl-diaminopimelate desuccinylase	Vibrio fischeri ES114	337/377 (89%),	358/377 (94%),	673	0.0	gnl|CDD|25776	99.7%	gi|39655034|pdb|1VGY|B	279/378 (73%),
54.	129	M16B	156VSmp	gi|58417186|emb|CAI28299.1| Insulinase (Peptidase family M16)	Hypothetical zinc protease	Ehrlichia ruminantium	34/116 (29%),	54/116 (46%),	33.9	1.9	XXXXXXXXXXXXXXX	XXXXXXXXXXXXXXX	XXXXXXXXXXXXXX	XXXXXXXXXXXXXX
55.	497	M32	156VSmp	gi|59480165|gb|AAW85952.1| Carboxypeptidase Taq (M32) metallopeptidase	thermostable carboxypeptidase 1	Vibrio fischeri ES114	442/494 (89%),	463/494 (93%),	884	0.0	gnl|CDD|11996	99.4%	gi|67463771|pdb|1WGZ|C	276/505 (54%),
56.	927	M16A	157VSm	gi|59712416|ref|YP_2	insulin-	Vibrio fischeri	786/910	843/910	1583	0.0	gnl|CDD|	95.8%	gi|67464134|	425/875

			p	05192.1| Secreted/periplasmic Zn-dependent peptidases	degrading enzyme	ES114	(86%),	(92%),			10753		pdb|1Q2L|A	(48%),
57.	1134	M43B	159VSmp	gi|82492138|gb|ABB77935.1| Zinc-dependent metalloprotease	hypothetical protein	Halophage AAJ-2005	33/85 (38%),	49/85 (57%),	55.8	1e-05	gnl|CDD|29068	72.6%	XXXXXXX XXXXXXX	XXXXXX XXXXXX XX
58.	512	S16	160VSmp	gi|84393426|ref|ZP_00992183.1| AAA-superfamily associated with proteolysis etc	ComM-related protein	Vibrio parahaemolyticus RIMD 2210633	400/508 (78%),	443/508 (87%),	775	0.0	gnl|CDD|10476	100.0%	gi|13399758|pdb|1G4B|F	27/59 (45%),
59.	382	M20A	166VSmp	gi|59712521|ref|YP_205297.1| zinc metallo Glutamate carboxypeptidases	succinyl-diaminopimelate desuccinylase	Vibrio fischeri ES114	337/377 (89%),	358/377 (94%),	673	0.0	gnl|CDD|10494	98.8%	gi|39655034|pdb|1VGY|B	279/378 (73%),
60.	485	M48B	166VSmp	gi|59712529|ref|YP_205305.1| Tetratricopeptide repeat domain present	zinc metalloprotease	Vibrio fischeri ES114	404/458 (88%),	433/458 (94%),	715	0.0	gnl|CDD|13919	92.8%	gi|71041595|pdb|1TI8|A	28/59 (47%),
61.	473	U32	171VSmp	gi|59712597|ref|YP_205373.1|	peptidase family U32	Vibrio fischeri ES114	441/462 (95%),	452/462 (97%),	886	0.0	gnl|CDD|23156	99.6%	gi|82407272|pdb|1OB5|E	47/108 (43%),
62.	328	T2	172VSmp	gi|76258751|ref|ZP_00766405.1|	Peptidase T2, asparaginase 2	Photobacterium profundum 3TCK	94/262 (35%),	147/262 (56%),	136	1e-30	gnl|CDD|30151	100.0%	gi|51247536|pdb|1T3M|D	78/152 (51%),
63.	220	C56	173VSmp	gi|63254267|gb|AAY35363.1| ThiJ/PfpI	Putative intracellular protease/amidase	Pseudomonas syringae pv. syringae	109/212 (51%),	152/212 (71%),	234	2e-60	CDD|10562	98.2%	gi|30750053|pdb|1OY1|D	142/212 (66%)
64.	488	U62	176VSmp	Putative modulator of DNA gyrase	putative TldD, Zn-dependent protease	Vibrio fischeri ES114	455/481 (94%),	473/481 (98%),	843	0.0	gnl|CDD|8039	100.0%	gi|55670448|pdb|1VL4|B	50%
65.	471	U62	176VSmp	Predicted Zn-dependent proteases and their inactivated homologs	PmbA protein	Vibrio splendidus 12B01	354/447 (79%),	401/447 (89%),	695	0.0	gnl|CDD|10186	99.6%	gi|56966588|pdb|1VPB|A	218/431 (50%),
66.	535	U62	176VSmp	Large exoproteins involved in heme utilization or adhesion	MSHA biogenesis protein MshQ	Vibrio splendidus 12B01	144/499 (28%),	221/499 (44%),	143	2e-32	-	-		
67.	786	S16	178VSmp	Lon protease (S16) C-terminal proteolytic domain	ATP-dependent protease La	Vibrio fischeri ES114	666/776 (85%),	715/776 (92%),	1332	0.0	gnl|CDD|10791	88.7%	gi|73535793|pdb|1Z0W|A	53%
68.	329	S1A	179VSmp	Trypsin-like serine protease	elastase 2 precursor	Vibrio fischeri ES114	172/268 (64%),	207/268 (77%),	335	1e-90	gnl|CDD|25284	99.6%	gi|229744|pdb|1CHG|	106/230 (46%),

69.	590	M23B	182VSmp	Lysin domain, involved in bacterial cell wall degradation	N-acetylmuramoyl-L-alanine amidase	Vibrio fischeri ES114	493/579 (85%),	535/579 (92%),	710	0.0	gnl|CDD|29017	100.0%		
70.		M44	184VSmp		Membrane-bound metallopeptidase						gnl|CDD|14075	63.8%		
71.	323	S33	185VSmp	Abhydrolase_1	hypothetical protein VF0216	Vibrio fischeri ES114	271/322 (84%),	296/322 (91%),	565	1e-159	gnl|CDD|23022	94.6%	gi|21465463|pdb|1GL7|G	19/29 (65%),
72.	856	S8A	185VSmp	involved in bacterial cell wall degradation	LysM repeat				395	3e-108				
73.	393	M17	186VSmp	Amidase	hypothetical protein VF1486	Vibrio fischeri ES114	308/388 (79%),	346/388 (89%),	599	8e-170				
74.	256	S51	191VSmp	Type 1 glutamine amidotransferase (GATase1)-like domain	peptidase E	Vibrio fischeri ES114	213/245 (86%),	231/245 (94%),	491	1e-137	gnl|CDD|28881	94.3%	gi|13096714|pdb|1FY2|A	128/208 (61%),
75.	391	S1C	194VSmp	PDZ domain of tryspin-like serine proteases, such as DegP/HtrA, which are oligomeric proteins involved in heat-shock response, chaperone function, and apoptosis	protease DegS precursor	Vibrio fischeri ES114	315/352 (89%),	333/352 (94%),	564	2e -159	gnl|CDD|10140	91.4%	gi|56966066|pdb|1TE0|B	221/321 (68%),
76.	455	S1C	194VSmp	Trypsin-like serine proteases, typically periplasmic, contain C-terminal PDZ domain [Posttranslational modification, protein turnover, chaperones]	endopeptidase DegP	Vibrio fischeri ES114	367/427 (85%),	406/427 (95%),	705	0.0	gnl|CDD|10140	91.6%	gi|20151111|pdb|1KY9|B	318/443 (71%),
77.		S33	197VSmp	alpha/beta hydrolase fold	lysophospholipase L2									
78.	639	M24B	198VSmp	X-Prolyl Aminopeptidase 2	Xaa-Pro aminopeptidase	Vibrio fischeri ES114	518/596 (86%),	556/596 (93%),	1095	0.0	gnl|CDD|9882	97.4%	gi|47168567|pdb|1PV9|B	113/217 (52%),
79.	922	M16B	198VSmp	Insulinase	zinc protease	Vibrio fischeri ES114	594/916 (64%),	740/916 (80%),	1189	0.0	gnl|CDD|24762	100.0%	gi|82407276|pdb|1SQP|A	109/229 (47%),

<table>
<tr><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></tr>
<tr><td>80.</td><td>407</td><td>M20B</td><td>199VSmp</td><td>Peptidase family M20/M25/M40</td><td>tripeptidase T</td><td>Vibrio fischeri ES114</td><td>341/367 (92%),</td><td>357/367 (97%),</td><td>652</td><td>0.0</td><td>gnl|CDD|11903</td><td>98.6%</td><td>gi|2780970|pdb|1CG2|D</td><td>162/350 (46%),</td></tr>
<tr><td>81.</td><td>495</td><td>M1</td><td>200VSmp</td><td>Peptidase_M1</td><td>membrane alanine aminopeptidase</td><td>Vibrio fischeri ES114</td><td>446/496 (89%),</td><td>466/496 (93%),</td><td>913</td><td>0.0</td><td>gnl|CDD|16956</td><td>99.0%</td><td>gi|71042175|pdb|1Z5H|B</td><td>178/445 (40%),</td></tr>
<tr><td>82.</td><td>1519</td><td>M1</td><td>200VSmp</td><td>hydrolysing acidic, basic or neutral N-terminal residues</td><td>membrane alanine aminopeptidase</td><td>Vibrio fischeri ES114</td><td>761/873 (87%),</td><td>812/873 (93%),</td><td>1519</td><td>0.0</td><td>gnl|CDD|10182</td><td>99.9%</td><td>gi|71042175|pdb|1Z5H|B</td><td>178/445 (40%),</td></tr>
<tr><td>83.</td><td>685</td><td>S41A</td><td>200VSmp</td><td>PDZ domain of C-terminal processing-, tail-specific-, and tricorn proteases, which function in posttranslational protein processing, maturation, and disassembly or degradation</td><td>Periplasmic protease, carboxy-terminal processing protease precursor</td><td>Vibrio fischeri</td><td>623/672 (92%),</td><td>649/672 (96%),</td><td>1228</td><td>0.0</td><td>gnl|CDD|26109</td><td>99.4%</td><td>gi|13096477|pdb|1FCF|A</td><td>171/374 (45%),</td></tr>
<tr><td>84.</td><td>432</td><td>M17</td><td>201VSmp</td><td>Remove N-terminal amino acid from a peptide or arylamide</td><td>aminopeptidase B, Cytosol aminopeptidase family</td><td>Vibrio fischeri ES114</td><td>374/428 (87%),</td><td>403/428 (94%),</td><td>747</td><td>0.0</td><td>gnl|CDD|7811</td><td>100.0%</td><td>gi|21730302|pdb|1GYT|L</td><td>173/304 (56%),</td></tr>
<tr><td>85.</td><td>372</td><td>M38</td><td>203VSmp</td><td>catalyzes the reversible interconversion of carbamoyl aspartate to dihydroorotate</td><td>dihydroorotase</td><td>Vibrio fischeri ES114</td><td>322/342 (94%),</td><td>334/342 (97%),</td><td>663</td><td>0.0</td><td>gnl|CDD|30037</td><td>100.0%</td><td>gi|66360534|pdb|1XGE|B</td><td>230/340 (67%),</td></tr>
<tr><td>86.</td><td>34</td><td>C15</td><td>204VSmp</td><td>cleaving pyroglutamate (pGlu) from the N-terminal end of specialized proteins</td><td>pyrrolidone-carboxylate peptidase</td><td>Vibrio fischeri ES114</td><td>27/31 (87%),</td><td>30/31 (96%),</td><td>53.1</td><td>3e-06</td><td>-</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td>87.</td><td>456</td><td>M50B</td><td>207VSmp</td><td>PDZ domain, presumably membrane-associated or integral membrane proteases, which may be involved in signalling and regulatory mechanisms</td><td>membrane metalloprotease</td><td>Vibrio fischeri ES114</td><td>377/452 (83%),</td><td>417/452 (92%),</td><td>768</td><td>0.0</td><td>gnl|CDD|29046</td><td>100.0%</td><td>gi|20151111|pdb|1KY9|B</td><td>52/95 (54%),</td></tr>
<tr><td>88.</td><td>354</td><td>M24A</td><td>207VSmp</td><td>catalyzes the removal of N-terminal amino acids from peptides and arylamides</td><td>methionine aminopeptidase</td><td>Vibrio fischeri ES114</td><td>252/275 (91%),</td><td>267/275 (97%),</td><td>546</td><td>6e-154</td><td>gnl|CDD|29971</td><td>100.0%</td><td>gi|9257170|pdb|4MAT|A</td><td>205/270 (75%),</td></tr>
</table>

89.	480	S8A	207VSmp	Subtilase family	Hypothetical extracellular protease	Photobacterium profundum SS9	390/480 (81%),	432/480 (90%),	753	0.0	gnl|CDD|25387	92.0%	gi|6573500|pdb|1DBI|A	143/261 (54%),
90.	238	M22	207VSmp	gi|59712319|ref|YP_205095.1|	glycoprotease protein	Vibrio fischeri ES114	193/233 (82%),	212/233 (90%),	384	2e-105	gnl|CDD|25625	63.5%	gi|55669534|pdb|1OKJ|D	165/231 (71%),
91.	390	S33	210VSmp	gi|91223152|ref|ZP_01258418.1| alpha/beta hydrolase fold	putative prolyl aminopeptidase	Vibrio alginolyticus 12G01	257/379 (67%),	303/379 (79%),	513	8e-144	gnl|CDD|23022	59.4%	gi|4389344|pdb|1AZW|B	41/82 (50%),
92.	2890	S8A	210VSmp	gi|27366005|ref|NP_761533.1| RTX toxin and related Ca2+-binding proteins	Autotransporter adhesin	Vibrio vulnificus CMCP6	765/1528 (50%),	1000/1528 (65%),	973	0.0	gnl|CDD|28913	98.0%	gi|24987447|pdb|1K7G|A	41/85 (48%),
93.	2908	G1	210VSmp	gi|6049492|gb|AAF02618.1|G-protein-coupled receptor proteolytic site domain	starry night protein	Drosophila melanogaster	196/790 (24%),	329/790 (41%),	137	5e-30	gnl|CDD|28913	98.0%	gi|999638|pdb|1SRP|	39/85 (45%),
94.	2890	S6	210VSmp	gi|59714345|ref|YP_207120.1|Von Willebrand factor type A (vWA) domain was originally found in the blood coagulation protein	iron-regulated protein FrpC	Vibrio fischeri ES114	731/1211 (60%),	921/1211 (76%),	1027	0.0	gnl|CDD|28913	98.0%	gi|24987254|pdb|1GO7|P	41/85 (48%),
95.	447	T1B	211VSmp	gi|59712885|ref|YP_205661.1| AAA-superfamily of ATPases associated with a wide variety of cellular activities, including membrane fusion, proteolysis, and DNA replication	ATP-dependent Hsl protease ATP-binding subunit HslU	Vibrio fischeri ES114	419/444 (94%),	430/444 (96%),	753	0.0	gnl|CDD|10938	100.0%	gi|13399760|pdb|1G4B|L	386/444 (86%),
96.	180	T1B	211VSmp	gi|36788015|emb|CAE17134.1|	ATP-dependent protease HslV (heat shock protein)	Photorhabdus luminescens subsp. laumondii TTO1	124/172 (72%),	144/172 (83%),	236	4e-61	gnl|CDD|30160	100.0%	gi|3114400|pdb|1NED|C	143/171 (83%),
97.	378	M14X	211VSmp	gi|59714036|ref|YP_206811.1| Predicted carboxypeptidase [Amino acid transport and metabolism	zinc-carboxypeptidase precursor	Vibrio fischeri ES114	345/374 (92%),	358/374 (95%),	730	0.0	gnl|CDD|12218	98.4%	gi|2781013|pdb|1NSA|	38/82 (46%),

98.	1230	C39	212VSmp	gi|82703906|ref|YP_413471.1| Peptidase family C39 mostly contains bacteriocin-processing endopeptidases from bacteria	Peptidase C39, bacteriocin processing	Nitrosospira multiformis ATCC 25196	299/1241 (24%),	502/1241 (40%),	263	3e-68	gnl|CDD|26072	61.8%	gi|2392399|pdb|1JSW|D	31/58 (53%),
99.	609	M3B	212VSmp	gi|59711846|ref|YP_204622.1| Peptidase_M3	oligoendopeptidase F	Vibrio fischeri ES114	479/592 (80%),	519/592 (87%),	926	0.0	gnl|CDD|10883	99.0%	gi|29726560|pdb|1N71|D	16/21 (76%),
100.	481	M23B	215VSmp	gi|59480839|gb|AAW86626.1|	cell wall endopeptidase, family M23/M37	Vibrio fischeri ES114	387/436 (88%),	408/436 (93%),	768	0.0	gnl|CDD|2090	90.4%	gi|88192196|pdb|2B13|B	55/93 (59%),
101.	303	S26A	216VSmp	gi|59712695|ref|YP_205471.1|	signal peptidase I	Vibrio fischeri ES114	282/300 (94%),	293/300 (97%),	577	2e-163	gnl|CDD|1022	100.0%	gi|51247605|pdb|1T7D|B	172/256 (67%),
102.	332	M23B	216VSmp	gi|59712675|ref|YP_205451.1|	lipoprotein NlpD	Vibrio fischeri ES114	264/319 (82%),	285/319 (89%),	387	3e-106	gnl|CDD|10607	94.6%	gi|88192196|pdb|2B13|B	52/109 (47%),
103.	334	M19	217VSmp	gi|75856936|ref|ZP_00764557.1| renal dipeptidase (rDP), membrane-bound glycoprotein hydrolyzing dipeptides	Zn-dependent dipeptidase, microsomal dipeptidase homolog	Vibrio sp. Ex25	301/329 (91%),	323/329 (98%),	659	0.0	gnl|CDD|12027	97.1%	gi|23200145|pdb|1ITU|B	93/187 (49%),
104.	881	M19	217VSmp	gi|71982905|ref|NP_001021076.1|	Hypothetical protein D2030.2a	Caenorhabditis elegans	62/258 (24%),	111/258 (43%),	36.6	5.9	-------	.------		
105.	525	U32	218VSmp	gi|90412956|ref|ZP_01220955.1|	putative collagenase family protease	Photobacterium profundum 3TCK	377/524 (71%),	434/524 (82%),	758	0.0	gnl|CDD|23156	30.4%	gi|60593939|pdb|1WWR|D	38/78 (48%),
106.	210	S24	218VSmp	gi|68544573|ref|ZP_00584207.1|	Helix-turn-helix motif:Peptidase S24, S26A and S26B	Shewanella baltica OS155	73/212 (34%),	113/212 (53%),	107	4e-22	gnl|CDD|11682	91.0%	gi|493805|pdb|1ADR|	35/62 (56%),
107.	180	U32	218VSmp	gi|90407496|ref|ZP_01215679.1|	ATP-dependent Zn protease	Psychromonas sp. CNPT3	48/142 (33%),	78/142 (54%),	65.1	1e-09	gnl|CDD|23842	95.7%	gi|61680474|pdb|1XZ0|C	20/37 (54%),
108.	382	M38	219VSmp	gi|75817876|ref|ZP_00748139.1|							gnl|CDD|30051	20.7%		
109.	795	S16	219VSmp	gi|59711405|ref|YP_204181.1|	ATP-dependent protease La	Vibrio fischeri ES114	751/784 (95%),	771/784 (98%),	1403	0.0	gnl|CDD|10339	99.9%	gi|40889711|pdb|1RR9|F	182/200 (91%),

110.	220	S14	219VSmp	gi|59711403|ref|YP_204179.1|	ATP-dependent Clp protease proteolytic subunit	Vibrio fischeri ES114	205/207 (99%),	207/207 (100%),	388	9e-107	gnl|CDD|23027	100.0%	gi|3318866|pdb|1TYF|N	177/193 (91%),
111.	872	M3B	221VSmp	gi|77977882|ref|ZP_00833320.1| VFDB virulence Factor	ATPases with chaperone activity, ATP-binding subunit	Yersinia intermedia ATCC 29909	79/259 (30%),	126/259 (48%),	105	1e-20	----	-----		
112.	409	M22	234VSmp	gi|59712856|ref|YP_205632.1|	O-sialoglycoprotein endopeptidase	Vibrio fischeri ES114	321/338 (94%),	329/338 (97%),	632	8e-180	gnl|CDD|10404	98.5%	gi|55669534|pdb|1OKJ|D	54/109 (49%),
113.	864	M13	236VSmp	gi|59711173|ref|YP_203949.1|	ClpB protein	Vibrio fischeri ES114	808/861 (93%),	842/861 (97%),	1541	0.0	gnl|CDD|10413	99.7%	gi|38492939|pdb|1QVR|C	629/855 (73%),