A novel strategy for the periplasmic production of heterologous proteins in *E. coli*

Dissertation

zur Erlangung des akademischen Grades

Doktor rerum naturalium (Dr.rer.nat.)

vorgelegt der

Naturwissenschaftlichen Fakultät I - Biowissenschaften

der Martin-Luther-Universität Halle-Wittenberg

Institut für Biochemie und Biotechnologie

von

Ajamaluddin Malik

geb. am 14. Mai 1976 in Siddharth Nagar, India

urn:nbn:de:gbv:3-000011925
[http://nbn-resolving.de/urn/resolver.pl?urn=nbn%3Ade%3Agbv%3A3-000011925]

Halle (Saale)

Gutachter:

1. Prof. Dr. Rainer Rudolph, Martin Luther University, Halle, Germany
2. Prof. Rudolf Glockshuber, ETH-Zurich, Switzerland
3. Prof. Dr. Robert Seckler, Potsdam University, Germany

verteidigt am 29.05.2007

This work is dedicated to

My parents, who encouraged me

My teachers, who enabled me

Acknowledgement

A journey is easier when you travel together. This thesis is the result of nearly five years of work whereby I have been accompanied and supported by many people. It is my pleasure that I have now the opportunity to express my gratitude for all of them.

The first person I would like to thank is my supervisor, Prof. Dr. Rainer Rudolph. With whom I have been involved since March 2002. I could not have imagined having a better advisor and mentor for my Ph.D. work. His knowledge, deep perceptive and valued association throughout the execution of this study, made this work a wonderful experience. During these years I have known Prof. Rudolph as a sympathetic and principle-centered person. I owe him lots of gratitude for having shown me this way of research.

I would like to thank my promotor Dr. Brigitte Söhling who kept an eye on the progress of my work and was always available when I needed her advice. She has a sharp eye on details and possesses superb analytical skill, which proved instrumental in the success of this project.

I am highly indebted to Dr. Elisabeth Schwarz for her thoughtful advice and moral support. I sincerely acknowledge the co-operation of all the faculty members of this institute as well as my fellow colleagues for providing me an excellent work environment during the past years. I owe my special thanks to Ms. Christiane Harnisch for technical help.

I am thankful to Dr. A. Schierhorn, Halle, for mass spectrometry, and Dr. P. Rücknagel, Halle, their help rendered during N-terminal sequencing

I am grateful to the stiftung Industrieforschung, Germany, and Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) of Germany for providing grants to carry out research. I am also grateful to the Institute for Biochemistry and Biotechnology, Martin Luther University, for providing me an excellent work culture.

I express my heartfelt gratitude to my wife Huma, for her love and patience during the Ph.D. period. One of our best experiences that we had during in this period was the birth of our daughter, Sarah, who provided an additional and joyful dimension to our life.

Last, but not least, I feel a deep sense of gratitude for my father and mother who formed part of my vision and taught me the good things that really matter in life. Ofcourse, i am highly thankful to my brothers Jamaluddin, Hisamuddin and sister Zubaida for their doubtless trust and confidence in me. I am proud to be one of them.

Zusammenfassung

Die Einführung rekombinanter DNA-Technologie hat die Nutzung von Proteinen in der Diagnostik und Therapie extrem erweitert, denn viele dieser Proteine konnten bis zu diesem Zeitpunkt nur in sehr geringen Mengen aus natürlichen Quellen gewonnen werden. Nun können sie rekombinant hergestellt werden, sowohl in naturidentischer Form, aber auch als spezielle Varianten mit besonderen Eigenschaften (z.B. Insulin-Analoga). Die Produktion im Bakterium *Escherichia coli* ist wirtschaftlich sehr attraktiv, zudem ist der Organismus biochemisch und genetisch überaus gut charakterisiert. Allerdings ist die Gewinnung des Produktes in nativer Form trotz aller Kenntnis bislang immer noch eine Herausforderung.

Proteine, die in der Diagnostik und Therapie eingesetzt werden, sind vielfach sezernierte Proteine, wie z.B. Hormone, Wachstumsfaktoren, Antikörper etc. Sie enthalten in der Regel Cysteine, die Disulfidbrücken ausbilden. Die korrekte Ausbildung der Disulfidbrücken ist für die Struktur und biologische Aktivität essentiell. Werden die Proteine im Cytoplasma von *E. coli* produziert, fallen sie als Einschlusskörper (*inclusion bodies*) aus. Die *in vitro* Rückfaltung kann extrem aufwendig und schwierig sein. Durch Sekretion der Proteine ins bakterielle Periplasma ist eine oxidative Faltung der Proteine und damit die Ausbildung der nativen Form direkt möglich. Weitere Vorteile der Produktion im Periplasma sind, dass die Proteine weniger stark durch Proteasen angegriffen werden, und dass die Reinigung der Proteine erleichtert sein kann. Für die Translokation der Proteine ins Periplasma ist eine prokaryotische Signalsequenz erforderlich, ihre Gegenwart allein ist aber keine Garantie für den Transfer, denn die Sequenz des maturen Proteins hat ebenfalls einen Einfluss. Es kann daher vorteilhaft sein, das rekombinante Protein als Fusion mit einem bakteriellen periplasmatischen Protein zu exprimieren.

In der vorliegenden Arbeit wurden humanes Pepsinogen und humanes Proinsulin als Modellproteine verwendet, um ein periplasmatisches Expressionssystem für die Produktion disulfidverbrückter Proteine in nativer Form im bakteriellen Periplasma zu etablieren. Beide Proteine enthalten je drei Disulfidbrücken. Humanes Pepsinogen wurde zunächst mit drei verschiedenen bakteriellen Signalsequenzen (pelB, ompT und dsbA) fusioniert. Die Gene wurden mittels des T7 Promotors exprimiert, aber natives Pepsinogen konnte nicht im Periplasma

detektiert werden. Die Expression war sehr stark, so dass Pepsinogen unprozessiert im Cytosol verblieb. Dann wurde pelB-Pepsinogen in pTrc99a und in pBAD22 kloniert. Nach Expression ausgehend vom *trc* Promotor (pTrc99a) wurde Pepsinogen im Periplasma nachgewiesen (ca. 16 µg/L OD1). Jedoch waren die bestimmten Ausbeuten nicht parallel reproduzierbar, auch nach Testung verschiedenster Wirtsstämme und Kultivierungsbedingungen. Die Expression ausgehend vom *araBAD* Promotor war so gering, so dass keine signifikante Pepsin-Aktivität im Periplasma nachweisbar war.

In einem neuartigen Ansatz wurde humanes Pepsinogen an den C-terminus von Ecotin (*E. coli* trypsin inhibitor), eine homodimeres periplasmatisches Protein (16 kDa) fusioniert. Jede Ecotin-Untereinheit enthält eine Disulfidbrücke. Ecotin ist sehr stabil, und bleibt selbst bei Inkubation für 30 min bei 100 °C, oder pH 1,0 nativ. Die Ecotin-Pepsinogen-Fusion wurde in pTrc99a exprimiert und mit Hilfe der Ecotin-Signalsequenz ins Periplasma transloziert. Nach Extraktion der periplasmatischen Fraktion und säurekatalysierter Aktivierung wurde das Ecotin-Pepsinogen-Fusionsprotein zu Pepsin umgewandelt. Somit lag das Fusionsprotein im Periplasma in nativer Form vor. Nach Anzucht von *E. coli* in Schüttelkolben wurden ca. 100 µg Fusionsprotein /L OD1 produziert, dies entsprach etwa 70 µg Pepsinogen/L OD1. Aus Biomasse, die durch Anzucht von *E. coli* im Fermenter gewonnen wurde, konnte das Fusionsprotein in drei Schritten, durch Ni-NTA-Chromatographie, Ionenaustausch- und Gelfiltrationschromatographie, mit einer Ausbeute von 23% zur Homogenität gereinigt werden. Dies entsprach einer Ausbeute von 7,6 mg nativem gereinigtem Pepsinogen pro Liter Fermentationsmedium.

Um Pepsinogen im Periplasma-Extrakt zu bestimmen, wurde ein hochsensitives Nachweissystem benötigt, da die Mengen an Pepsinogen in den periplasmatischen Proben sehr gering waren. Im Rahmen dieser Arbeit wurde ein fluorimetrischer Assay für Pepsin und Pepsinogen entwickelt, bei dem eine Variante des grünfluoreszierenden Proteins, EGFP (enhanced green fluorescent protein) als Substrat diente. Nach saurer Denaturierung wies EGFP keine Fluoreszenz mehr auf. Die Fluoreszenz war jedoch durch nachfolgende Neutralisierung vollständig wiederherstellbar. In dem neu entwickelten Proteolyse-Assay wurde säuredenaturiertes EGFP mit Pepsin oder aktiviertem Pepsinogen gespalten. Nach Neutralisierung wurde die Menge an verbliebenem intakten EGFP anhand der Fluoreszenz bestimmt. Die Sensitivität des neuen Proteolyse-Assays war abhängig von der Inkubationstemperatur und Dauer der Proteolyse. Wenn die Spaltung für

drei Stunden bei 37 °C durchgeführt wurde, waren selbst 50pg Pepsin ausreichend für ein signifikantes Signal. Unter Standard-Bedingungen, bei einer Inkubationszeit von 10 min bei 20 °C, lag die Sensitivität des Assays bei 0-30 ng Pepsin bzw. aktiviertem Pepsinogen. Für Pepsin aus Schwein betrug die spezifische Umsatzrate 38 ± 6,7 ng EGFP x ng^{-1} Pepsin x min^{-1}. Für säureaktiviertes Pepsinogen wurde eine vergleichbare Umsatzrate (37,2 ± 5,2 ng EGFP x ng^{-1} aktiviertes Pepsinogen x min^{-1}) bestimmt. Die Pepsin-katalysierte Spaltung von EGFP zeigte eine typische Michaelis-Menten Kinetik.

Um die Anwendbarkeit von Ecotin als ein periplasmatisches Fusionsprotein zu evaluieren, wurde ein zweites humanes Protein aus einer gänzlich anderen Familie ausgewählt. Humanes Proinsulin enthält drei nicht-konsekutive Disulfidbrücken. Die Sequenz für Proinsulin wurde genetisch mit der Sequenz für Ecotin fusioniert. Das Fusionprotein wurde in pTrc99a kloniert und in *E. coli* BL21(DE3)Gold exprimiert. Die Parameter für die Proteinproduktion wurden optimiert. Nach Anzucht bei hoher Zelldichte im Fermenter wurde eine Ausbeute von 153 mg Ecotin-Proinsulin pro Liter Fermentationsmedium bestimmt. Aus der Biomasse wurde die periplasmatische Fraktion extrahiert, und das Fusionsprotein wurde anhand eines neu etablierten Reinigungsverfahrens durch Affinitätschromatographie in einem Schritt gereinigt. Da Ecotin ein Trypsin-Inhibitor ist, kann immobilisiertes Trypsinogen oder eine enzymatisch inaktive Trypsin-Variante als selektive Affinitätsmatrix verwendet werden. Natives humanes Proinsulin wurde zur Homogenität gereinigt, durch ELISA quantifiziert und durch Massenspektrometrie charakterisiert. Um den Effekt periplasmatischer Proteasen auf die Ausbeute an nativem Proinsulin zu untersuchen, wurde die Produktion von nativem Proinsulin in einem Wildtyp-Stamm, *E. coli* BL21(DE3)Gold und in einem Protease-defizienten Stamm, *E. coli* SF120, untersucht. Nach Anzucht im Schüttelkolben war die spezifische Ausbeute an Ecotin-Proinsulin in *E. coli* SF120 drei bis vierfach höher als im Wildtypstamm. Das hier beschriebene periplasmatische Fuisonprotein auf der Basis von *E. coli* Ecotin ist ein neuartiges, effizientes biochemisches Verfahren, um rekombinante Proteine effizient im bakteriellen Periplasma zu produzieren.

TABLE OF CONTENTS

Chapters **Page no.**

1 ABSTRACT……..……..…….…. **1**

2 INTRODUCTION…….……..... **4**

2.1 Recombinant protein expression in *E. coli*……..**4**

2.1.1 Occurrence and significance of disulfide bonds…….....5

2.1.2 Chemistry of disulfide bond formation……. 6

2.1.3 Choice of cellular compartment for protein expression…… 6

2.2 Periplasmic expression of heterologous proteins.....**8**

2.2.1 Anatomy and physiology of the periplasm……....8

2.2.2 Translocation of recombinant proteins…….……9

2.2.3 Targeting signals……... 9

2.2.4 Properties of the mature protein……..….. 11

2.2.5 Pathways of translocation……..….11

2.2.6 Protein folding in the periplasm…… 13

2.2.6.1 Disulfide bond formation……..… 14

2.2.6.2 Disulfide bond isomerization ……...… 15

2.2.6.3 Peptidyl-prolyl cis/trans isomerases…….……..……..17

2.2.6.4 The periplasmic chaperones…….…… 18

2.2.6.5 Proteolysis…….……..…….. 18

2.3 Periplasmic fusion proteins…….. **19**

2.3.1 Maltose-binding protein (MBP)……..... 19

2.3.2 Staphylococcal protein A (SpA)……...… 20

2.3.3 Streptococcal protein G (SpG)……...……..…… 20

2.3.4 Cellulose binding domain (CBD)……..……21

2.3.5 Disulfide bond oxidoreductase (DsbA)…….……21

2.3.6 Barnase…….…..21

2.4 Strategies for optimal expression……..… **22**

2.4.1 Molecular level……..…22

2.4.2 *E. coli* host strain……..…… 23

2.4.3 Stress response…….…… 23

2.4.4 Fermentation optimization…….24

3 OBJECTIVE……....**25**

4 RESULTS AND DISCUSSION……...…….. **26**

4.1 Development of a fluorogenic assay for pepsin……..… **26**

4.1.1 Evaluation of various pepsin assays……..…26
4.1.2 Use of EGFP as a substrate……... 28
4.1.3 Reversible unfolding of EGFP…….. 28
4.1.4 Sensitivity of the EGFP assay……...… 30
4.1.5 Kinetics of proteolysis……...……31
4.2 Periplasmic expression of human pepsinogen…… **34**
4.2.1 Cloning strategy for periplasmic expression of pepsinogen…… 34
4.2.2 Expression of pelB-pepsinogen, ompT-pepsinogen and dsbA-pepsinogen….. 35
4.2.3 Expression of pepsinogen upon fusion to *E. coli* ecotin……... 37
4.2.4 Large scale expression and purification of ecotin-pepsinogen…….…….38
4.2.5 Characterization of ecotin-pepsinogen…….……40
4.2.6 Future prospects of the ecotin fusion system……..……...……..……. 41
4.3 Periplasmic expression of human proinsulin with ecotin fusion……...**42**
4.3.1 Cloning of ecotin-proinsulin…….……42
4.3.2 Expression, purification and characterization of ecotin-proinsulin…..……..42
4.3.3 Optimization of ecotin-proinsulin expression…….…….. 45
4.3.4 Extraction of ecotin-proinsulin at large scale……47
4.3.5 Affinity purification of ecotin-proinsulin…….……47
4.3.6 Characterization of ecotin-proinsulin…….……... 49
4.3.7 Production of ecotin-proinsulin in the protease deficient strain……..……..49
5 REFERENCES……...…….…… **51**
6 ABBREVIATIONS…...……**64**
7 PUBLISHED AND SUBMITTED MANUSCRIPTS AND PATENTS……..……**66**
7.1 Malik, A., Rudolph, R., Söhling, B. (2005) Use of enhanced green fluorescent protein to determine pepsin at high sensitivity. *Analytical Biochemistry*, 340, 252–258…….…….…. 67
7.2 Malik, A., Rudolph, R., Söhling, B. (2006) A novel fusion protein system for the production of native human pepsinogen in the bacterial periplasm. Prot. Expr. Purif., 47, 662–671…….……....74
7.3 Malik, A., Jenzsch, M., Lübbert, A., Rudolph, R., Söhling, B.(2006) Periplasmic production of native human proinsulin as a fusion to *E. coli* ecotin (under revision-Prot. Expr.Purif)……...…….……84
7.4 Malik, A., Söhling, B. & Rudolph, R. (2005) "Periplasmatische Produktion therapeutisch relevanter Proteine durch Fusion an *Escherichia coli* Ecotin " Deutsche Patent- und Markenamt (Patent pending-10 2005 050 732.8-41)……....118
8 APPENDIX……...……..……....…… **150**

1 Abstract

Since the advent of recombinant DNA technology, application of proteins in pharmaceuticals was remarkably changed. All these proteins can not be obtained from natural sources because many of them are present in extremely low amounts. Genetically engineered proteins having special advantages (e.g. Insulin analogs) are as such artificial molecules and can therefore only be obtained recombinantly. *Escherichia coli* offer a means for the rapid and economical production of recombinant proteins. These advantages are coupled with a wealth of biochemical and genetic knowledge. Although significant progress has been made in improvement of transcription, translation and secretion, obtaining the product in a soluble and bioactive form is still a major challenge.

Many naturally secreted proteins such as hormones, growth factors, antibodies etc. are used for diagnostic and therapeutic applications. In general, if secreted proteins contain two or more cysteines then they form disulfide bonds that are essential for structure formation and function. Production of these proteins in the cytoplasm of *E. coli* usually gives inclusion bodies due to a reducing environment. *In vitro* oxidative refolding can be quite difficult. Secretion of such proteins into the periplasm of *E. coli* provides a better chance of oxidative folding due to the presence of oxidative folding and disulfide isomerization machinery. Besides the formation of correct disulfide bonds, production in the periplasm can also reduce the proteolysis and can ease the purification. For the translocation of proteins to the periplasm, a prokaryotic signal peptide is required, but the presence of this signal sequence does not always ensure efficient protein translocation. The sequence next to the signal peptide cleavage site in the mature part of the protein and other region of mature part play an important role in translocation. If this is the case, fusion to a full length periplasmic protein that is well produced and properly folded is more promising.

In this study, human pepsinogen and human proinsulin were used as model proteins to establish an expression system for the production of disulfide bonded proteins in native form in the periplasm of *E. coli*. Both, pepsinogen and proinsulin contain three disulfide bonds. For the production of pepsinogen, three different signal sequences (pelB, ompT and dsbA) were fused to its N-terminus for translocation. The genes were expressed from the T7 promoter in pET

vectors, but no pepsin activity could be determined in the periplasm. The expression level was very high, so that pepsinogen remained in the cytosol along with the signal sequence. Next, pelB-pepsinogen was cloned into pTrc99a and pBAD22 to replace the T7 promoter by the hybrid *trc* and by the weak *ara*BAD promoter, respectively. Using the *trc* promoter, about 16 µg pepsinogen per liter OD1 was determined in the periplasm. However, production of pepsinogen was not reproducible even though different strains and culture conditions were tested. In case of the *ara*BAD promoter, expression level was very poor and no significant pepsin activity could be determined.

As a new approach, human pepsinogen was fused to the C-terminus of ecotin, *E. coli* trypsin inhibitor, which is a homodimeric periplasmic protein (16 kDa). Each subunit contains one disulfide bond. It is a highly stable protein and withstands even 100 °C and pH 1.0 for 30 min. The ecotin-pepsinogen fusion was expressed in pTrc99a and was translocated into the periplasm with the help of the ecotin signal peptide. When the periplasmic extract was acidified, the ecotin-pepsinogen fusion was converted into pepsin, indicating that pepsinogen was produced in its native form. In shake flask experiments, the amount of native ecotin-pepsinogen present in the periplasm was 100 µg per liter OD1 that corresponds to 70 µg pepsinogen. After large scale cultivation, the native fusion protein was purified to homogeneity in three step of purification, Ni-NTA, ion exchange and gel filtration chromatography with a yield of 23%. From 30 g wet biomass, 5.2 mg ecotin-pepsinogen corresponding to 3.6 mg pepsinogen was obtained. This corresponded to 7.6 mg native pure pepsinogen per liter fermentation broth.

To identify and quantify pepsinogen in periplasmic samples, a highly sensitive assay method was needed due to the low amount of protein present in the periplasmic samples. A fluorometric assay for pepsin and pepsinogen was developed using enhanced green fluorescent protein (EGFP) as a substrate. Acid denaturation of EGFP resulted in a complete loss of fluorescence that was completely reversible on neutralization. In the proteolytic assay procedure, acid-denatured EGFP was digested by pepsin or activated pepsinogen. After neutralization, the remaining amount of undigested EGFP refolded and was quantified by fluorescence. The sensitivity of the proteolytic assay was dependent on the incubation time and temperature. If digestion of EGFP was done for 3 hours at 37 °C, even 50 pg pepsin were

sufficient to give a reasonable signal. Under standard digestion conditions at 20 °C for 10 min, the sensitivity of pepsin or activated pepsinogen was in the range of 0-30 ng. Using porcine pepsin, the specific digestion rate of EGFP under standard condition was 38 ± 6.7 ng EGFP ng^{-1} pepsin min^{-1}. Acid treated, activated porcine pepsinogen revealed a similar specific digestion rate (37.2 ± 5.2 ng EGFP ng^{-1} activated pepsinogen min^{-1}). The pepsin-catalyzed EGFP digestion showed typical Michaelis–Menten kinetics.

To evaluate the applicability of ecotin as a periplasmic fusion tag, a second human protein from a diverse family, proinsulin, was chosen. Proinsulin contains three non-consecutive disulfide bonds. It was genetically fused to the C-terminus of ecotin. The ecotin-proinsulin fusion was cloned downstream of the *trc* promoter in pTrc99a and the fusion protein was produced in *E. coli* BL21(DE3)Gold. Parameters were optimized for the improvement of ecotin-proinsulin production. In high cell density cultivation, 153 mg ecotin-proinsulin per liter broth was produced. Downstream processing was done in one step using a newly established affinity purification method. Since ecotin is a trypsin inhibitor, trypsinogen or inactive trypsin variants immobilized to a column can serve as a highly selective affinity material. Native human proinsulin was purified to homogeneity, estimated by ELISA and characterized by mass spectrometry. To evaluate the effect of proteolysis in the periplasm of *E. coli,* the amount of ecotin-proinsulin was determined in a wild-type strain, *E. coli* BL21(DE3)Gold, and in a strain deficient in several periplasmic protease, *E. coli* SF120. At the shake flask level, the specific yield of ecotin-proinsulin was 3-4 fold higher in *E. coli* SF120 than in a wild-type strain, *E. coli* BL21(DE3)Gold. In summary, the ecotin fusion protein system is a novel and useful tool to efficiently produce recombinant proteins in the bacterial periplasm.

2 Introduction

2.1 Recombinant protein expression in *E. coli*

Proteins either for therapeutic, industrial or basic research purposes are needed in large quantity. In nature, many of them are present in extremely low amounts, making it difficult or nearly impossible to maintain supply according to demand. However, recombinant DNA technology has provided a tool to produce proteins in bulk quantities. Recombinant proteins have two distinct advantages over traditional extraction of proteins from natural sources: there is virtually an unlimited supply of recombinant proteins, and recombinant proteins are identical to natural proteins (Chance*, et al.*, 1981). Heterologous proteins must be expressed at high levels and subsequently purified for either purpose, thus the most important parameter is the abundance of the protein at the time of harvesting.

E. coli remains a valuable host for the expression of heterologous proteins, and for many purposes is the best host (Pines and Inouye, 1999). The advanced knowledge concerning the genetics and physiology of *E. coli* has accounted for preferential use of *E. coli* as a host. Additional advantages of *E. coli* based expression systems include its rapid generation of biomass, low-cost culture conditions, and the availability of an increasingly large number of cloning vectors and mutant host strains (Baneyx, 1999). However, in spite of the extensive knowledge on the genetics and molecular biology of *E. coli*, not every gene can be expressed efficiently in this organism. This may be due to the unique and subtle structural features of the gene sequence, the stability and translational efficiency of mRNA, major differences in codon usage, protein folding, degradation of the protein by host cell proteases, and the potential toxicity of the protein to the host. Fortunately, some empirical “rules” that can guide the design of expression systems and limit the unpredictability of this operation in *E. coli* have emerged. The major drawbacks of *E. coli* as an expression system include the inability to perform many of the post-translational modifications found in eukaryotic proteins and lack of secretion mechanisms for the efficient release of recombinant proteins into the periplasm or culture medium. On the other hand, many eukaryotic proteins retain their full biological activity in a nonglycosylated form and therefore can be produced in *E. coli* (Fuh*, et al.*, 1990, Liang*, et al.*, 1985).

2.1.1 Occurrence and significance of disulfide bonds

Proteins containing stable disulfide bonds are rarely found in the cytoplasm except certain archaea (Mallick*, et al.*, 2002). In bacteria, they are usually located in extracytoplasmic compartments or secreted into the medium. In eukaryotic cells, they are present in compartments such as the endoplasmic reticulum and the plasma membrane, or they are secreted into the external milieu. Proteins that contain disulfide bonds can be divided into two classes: those in which the cysteine-cysteine linkage is a stable part of their final folded structure and those in which pairs of cysteines alternate between the reduced and oxidized states. For the first class, the disulfide bond may contribute to the folding pathway of the protein e.g. BPTI, and to the stability of its native state, e.g. ß-lactamase (Creighton, 1997b). For the second, the oxidative-reductive cycling of the disulfide bond may be central to a protein's activity as an enzyme, e.g. certain ribonucleotide reductases, or may be involved in a protein's activation and deactivation, e.g., OxyR (Zheng*, et al.*, 1998). The formation of a disulfide bond stabilizes a protein with about 4 kcal/mol per disulfide bond, mainly by decreasing the conformational entropy of the denatured state (Clarke and Fersht, 1993, Pantoliano*, et al.*, 1987, Shaw and Bott, 1996). Failure to form proper disulfide bonds is likely to cause protein misfolding, leading to aggregation and degradation by proteases. In this study, human pepsinogen and human proinsulin were used as model proteins to establish a platform for the periplasmic production of disulfide bonded proteins. Both proteins contain three disulfide bonds that are essential for proper folding and activity.

Disulfide bond formation in the cytoplasm of eukaryotes is largely prevented by the reduced environment, with a ratio of GSH/GSSG estimated to be in between 30:1 and 100:1 (Freedman, 1989). In the endoplasmic reticulum, the redox environment is "buffered" at a ratio of 3:1 of GSH and GSSG (Hwang*, et al.*, 1992). Reducing balance in the cytoplasm of *E. coli* is established with the glutaredoxins, thioredoxin and thioredoxin reductase system (Aslund*, et al.*, 1994, Carmel-Harel and Storz, 2000, Russel, 1995). Disulfide bond catalysis is slightly more complicated in the periplasm of *E. coli* than in the endoplasmic reticulum. The periplasm is separated from the extracellular environment by a porous membrane which allows passive diffusion of small molecules. Therefore, variations in the medium composition do not allow the

existence of a defined redox potential. The oxidizing environment in the periplasm is maintained by the presence of Dsb family of proteins (Bardwell*, et al.*, 1993).

2.1.2 Chemistry of disulfide bond formation

Disulfide bonds can form spontaneously in the presence of molecular oxygen. However, air oxidation is a rather slow process: *In vitro*, it can take several hours or even days to allow the formation of all the native disulfide bonds present in a protein. In contrast, disulfide bond formation *in vivo* occurs within minutes or even seconds after synthesis. For instance, the refolding of RNAse A, a protein with four disulfide bonds, takes several hours *in vitro* but less than 2 min *in vivo* (Sevier and Kaiser, 2002). A thiol–redox reaction between a molecule A containing two reduced cysteine residues and a molecule B containing one disulfide bond can be seen as the transfer of two electrons from A to B or as the transfer of a disulfide bond from B to A. The rate of the reaction depends on the accessibility of the reactive groups, on the difference of the redox potential between the redox partners and on the probability of the sulfur atoms to come within the distance required for thiol/disulfide exchange (Englander and Kallenbach, 1983). The oxidation of thiols requires an appropriate electron acceptor, and the reduction of a disulfide requires an electron donor. *In vivo*, the ultimate electron acceptor for thiol oxidation is usually molecular oxygen whereas under anaerobic conditions usually fumarate is a terminal electron acceptor. The ultimate source of electrons for disulfide reduction is usually NADPH.

2.1.3 Choice of cellular compartment for protein expression

Proteins have been successfully produced in the cytoplasm and periplasmic space of *E. coli.* In general proteins are more easily produced in the cytoplasm, but often aggregate into insoluble inclusion bodies. As soon as a nascent polypeptides chain emerges from the ribosome, it faces crowding of the cytoplasm. Exposed hydrophobic surfaces of polypeptides are prone to aggregation. To prevent this fatal process, cells possess a complex machinery of molecular chaperones that helps nascent proteins to fold. Another obstacle for disulfide bonded proteins is the reducing environment of the cytosol. When proteins are exported to more oxidizing environment of the periplasmic space, the chances of correct disulfide bond formation are high. Obtaining efficient translocation of a recombinant protein across the cytoplasmic membrane is

often a limitation. The decision to target the expressed protein to a specific cellular compartment rests on balancing the advantages and disadvantages of each compartment summarised in table 1.

Table 1: Summary of merits and demerits of different compartments[a] for gene expression in *E. coli*.

Compartment	Advantages	Disadvantages
Cytoplasm	Higher protein yields **As inclusion bodies:** - facile isolation - protection from proteolysis - inactive protein (cannot harm host)	**As inclusion bodies:** - refolding can be difficult - refolding yield might be low - high cost of folding additives **As soluble protein:** - S-S bond formation disfavoured - presence of N-terminal methionine - extensive proteolysis - purification is more complex
Periplasm	S-S bond formation favoured Purification is simpler Less extensive proteolysis N-terminal methionine processed Outer-membrane gives access to ~1 kDa molecules, suitable for screeing or folding additive purposes	Lower protein yields Signal peptide does not always work Inclusion bodies may form
Medium	Least extensive proteolysis Simpler purification Improved folding Processed N-terminal methionine	Usually no protein secretion Protein dilution

[a] Protein expression localized in inner membrane or cell surface are not mentioned in the table because to date they are not used for high-level gene expression.

2.2 Periplasmic expression of heterologous proteins

2.2.1 Anatomy and physiology of the periplasm

The periplasmic space lies between the inner and outer membrane of gram-negative bacteria, consisting of several distinct microenvironments created by two boundary membranes and the lysozyme sensitive peptidoglycan layer (Fig. 1). Each layer is ~7.5 nm in thickness (Hobot, *et al.*, 1984, Leduc, *et al.*, 1985). The peptidoglycan layer is located in the periplasmic space and serves as cytoskeleton that contributes to cell shape and prevents cells from lysing in dilute environments.

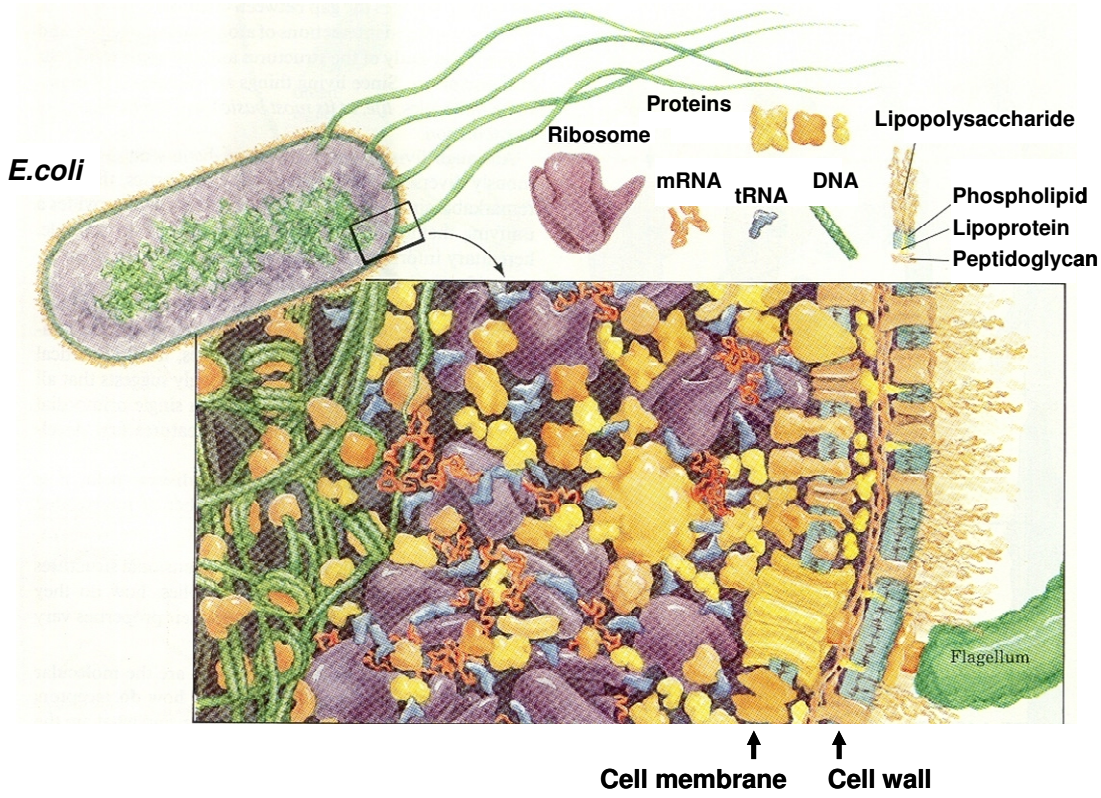


Fig. 1. The simulated cross-section of an *E. coli* cell magnified around one million fold (Voet and Voet, 1995; modified)
The periplasm shown on right hand side of the drawing is surrounded by a multilayer outer cell wall and inner cell membrane. Exterior surface of cell wall is decorated with lipopolysaccharides. The cytoplasm, which occupies the middle region of the drawing, is predominantly filled with ribosomes engaged in protein synthesis. The left side of drawing contains a dense tangle of DNA in complex with specific proteins.

Under normal physiological conditions the inner periplasmic space is approximately 4 nm thick in cross section and has content similar in density to that of the cytoplasm (Dubochet, *et al.*,

1983). Physiological and electron microscopic measurements show that the periplasmic space of *E. coli* represents ~20-40% of the total cell volume in normal growth media. Since the cytoplasmic membrane is permeable to water, the periplasm and cytoplasm are iso-osmolar (Stock*, et al.*, 1977, Vanwielink and Duine, 1990). Periplasmic polysaccharides and other small molecules create a significant Donnan equilibrium across the outer membrane and serve to buffer the cell from changing osmotic and ionic environments. This helps to preserve the more constant internal environment needed for cell growth and viability (Stock*, et al.*, 1977). Remarkably, as the periplasm is devoid of ATP, all the protein folding, trafficking and degradation takes place in the absence of obvious energy source (Mogensen and Otzen, 2005).

2.2.2 Translocation of recombinant proteins

Several pathways have been described for the export of proteins in *E. coli* (Fekkes and Driessen, 1999). Of these, the most widely used and conserved system is the general secretory (Sec) pathway. Periplasmic as well as inner and outer membrane proteins are recognized and targeted to the Sec pathway by short, cleavable, N-terminal signal sequences (Fekkes and Driessen, 1999). Protein transport occurring across the cytoplasmic membrane is usually classified according to its temporal relationship to protein synthesis as occurring co-translationally or post-translationally. Co-translational translocation requires a specialized membrane targeting device to guarantee simultaneous synthesis and transport, while a post-translational mode of protein transport requires molecular chaperones to prevent premature folding of the completed polypeptide chain after its release from the ribosome (Muller*, et al.*, 2001).

2.2.3 Targeting signals

Precursor proteins are equipped with signals that are recognized by targeting factors to direct them to the translocation site. The signal sequence ranges in length from 18 to about 30 amino acid residues (Fig. 2). It is composed of three domains: the positively charged amino terminus (N region); the nonpolar, hydrophobic core region (H region); and the more polar cleavage region (C region) (von Heijne, 1985). The amino acid sequences of these domains are not well conserved, but their physicochemical properties are (Izard and Kendall, 1994).

N-domain: The first few amino acids of signal sequences are positively charged (Lys and Arg). They enhance the processing and translocation rates of a precursor protein but are not essential. Preproteins with signal sequences that carry a neutral or even negatively charged N region can be processed, albeit at reduced rates (Gennity*, et al.*, 1990).

H-domain: The H domain is the hydrophobic core of a signal sequence and varies in length from 7 to 15 amino acids. It is the most important part of the signal sequence; the translocation efficiency increases with the length and hydrophobicity of the H region (Chou and Kendall, 1990).

C-domain: The signal peptide cleavage site (C-domain) is the only part of the signal sequence that has some primary sequence specificity. The conserved residues located at positions -1 and -3 relative to the start of the mature part of the protein (Dalbey and Vonheijne, 1992). Usually, these residues have small neutral side chains, such as alanine, glycine, serine, and threonine, with a preference for alanine (Vonheijne, 1984).

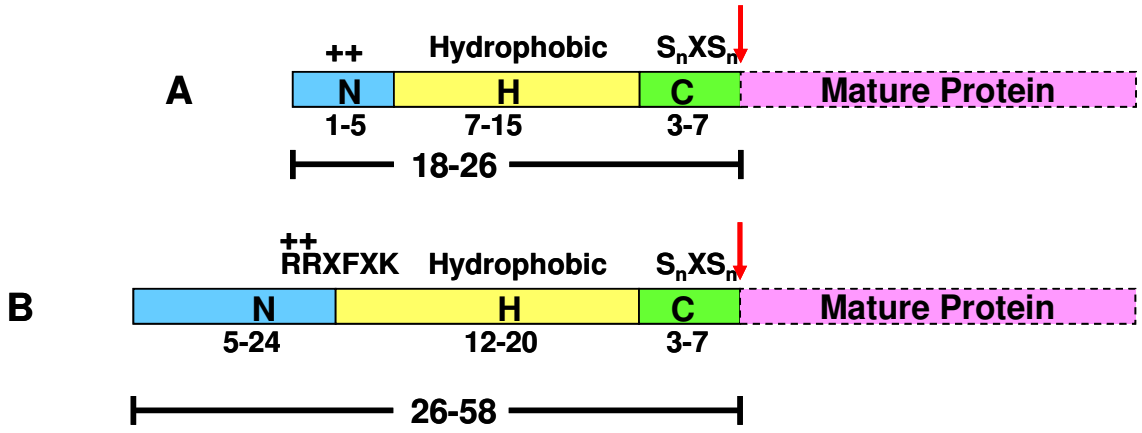


Fig. 2. Domain structure of the signal sequence of precursor proteins (Fekkes and Driessen, 1999; modified)

(A) Signal sequences of SRP- or SecB-dependent preproteins have a net positive charge in the N region (indicated by +), a hydrophobic H region, and a C region with the signal peptidase cleavage site (↓) preceded by the motif S_nXS_n, in which S_n stands for an amino acid with a small neutral side chain and X stands for any amino acid residue.

(B) Signal sequences of precursor proteins that are dependent on the twin-arginine route resemble normal signal sequences but have an extended N region and possess the RRXFXK motif, which straddles the H and C domains. For both types of signal sequences, the variation in length of the different regions and of the total signal sequence is indicated.

2.2.4 Properties of the mature protein

Targeting information may also be located within the mature part of the proteins. Certain proteins can be translocated even when they lack a signal sequence (Flower, *et al.*, 1994, Prinz, *et al.*, 1996). Other proteins are not translocated at all, when fused with a signal sequence (Lee, *et al.*, 1989, Malik, *et al.*, 2006). One reason for a lack of translocation is the presence of positively charged residues at the beginning of the mature domain (Kajava, *et al.*, 2000, Malik, *et al.*, 2006). Pepsinogen which was used in this study contains 13 basic residues in the prosegment (Sogawa, *et al.*, 1983).

2.2.5 Pathways of translocation

Three pathways can be used for secretion across the bacterial cytoplasmic membrane: the SecB-dependent pathway, the signal recognition particle (SRP), and the twin-arginine translocation (TAT) pathways (Mergulhao, *et al.*, 2005).

SecB-dependent pathway: The vast majority of secreted proteins are translocated by the SecB-dependent pathway. Ribosome-associated nascent chains of secreted proteins bind to the trigger factor, which is bound to the ribosomes (Maier, *et al.*, 2003). This association is maintained until the preprotein leaves the ribosome, thus preventing co-translational binding of the nascent chain to SRP components (Maier, *et al.*, 2003). SecA binding to the preprotein is facilitated by the signal peptide (Kebir and Kendall, 2002). Secreted proteins are kept in a translocation-competent state by the chaperone SecB (Luirink, *et al.*, 2005), which interacts with the mature region of the preprotein to prevent premature folding (Khokhlova and Nesmeianova, 2003). In the presence of the preprotein, SecB binds SecA (Woodbury, *et al.*, 2000). The complex of precursor-SecB and SecA is then bound to the SecY subunit of the SecYEG complex via SecA. ATP hydrolysis releases the preprotein from SecA into the translocation channel (Driessen, *et al.*, 1998). Binding of the preprotein to membrane-bound SecA results in the translocation of approximately 20 amino acids, and subsequent binding of ATP to SecA promotes SecA membrane insertion and translocation of additional 15-20 amino acids. Multiple rounds of SecA insertion and deinsertion promote protein translocation through the channel (de Keyzer, *et al.*, 2003). Proton-motive force (PMF) can complete translocation when the preprotein is halfway through the translocase, even

in the absence of SecA (Nishiyama, *et al.*, 1999). The translation of SecA is regulated by SecM, which monitors the secretion proficiency of *E. coli* (Oliver, *et al.*, 1998, Sarker, *et al.*, 2000).

SRP pathway: Signal recognition particle (SRP) is a ribonucleoprotein whose components, the 54 kDa protein (Ffh) and 4.5S RNA, are widely conserved across all domains of life (Luirink and Dobberstein, 1994). Proteins exported by the SRP pathway contain highly hydrophobic signal sequences (Lee and Bernstein, 2001). It appears that the SRP pathway is generally used for the assembly of the membrane proteins into the cytoplasmic membrane. Recently, however, evidence has been presented that some proteins (e.g. DsbA) with cleavable signal sequences are co-translationally exported in a SRP-dependent process to the periplasm (Schierle, *et al.*, 2003). An artificially increase in the hydrophobicity of other signal sequences will target them to this pathway (Bowers, *et al.*, 2003, Schierle, *et al.*, 2003). SRP binds the ribosome at a site that overlaps the binding site of a trigger factor (Gu, *et al.*, 2003). Depending upon the characteristics of the nascent peptide, the binding of either SRP or a trigger factor is stabilized, thus determining whether the peptide is targeted co-translationally via the SRP pathway, or post-translationally by the SecB pathway (Gu, *et al.*, 2003). FtsY is an *E. coli* SRP receptor, found both in the cytoplasm and at the membrane (de Leeuw, *et al.*, 1997). FtsY interacts with ribosomal nascent chain-SRP complexes (Herskovits, *et al.*, 2000). Upon interaction with membrane lipids, the GTPase activities of FtsY and Ffh are stimulated, thus releasing the nascent chain to the translocation site (Nagai, *et al.*, 2003). Insertion of transmembrane segments can occur in the absence of SecA (Scotti, *et al.*, 1999) while translocation of large periplasmic loops is SecA-dependent (Qi and Bernstein, 1999).

TAT pathway: The twin-arginine translocation (TAT) system is a Sec-independent pathway (Santini, *et al.*, 1998). The signal sequence contains two consecutive and highly conserved arginine residues (Fig. 2). The TAT pathway is capable of transporting folded proteins across the inner membrane using the transmembrane PMF (de Leeuw, *et al.*, 2002, Stanley, *et al.*, 2000). The main components of this translocation system are TatA, B, C, D and E (Mergulhao, *et al.*, 2005) but their function is not yet fully established. TatA has been proposed to form the transport channel (Palmer and Berks, 2003). TatB and TatC are proposed to form a 1:1 complex that may provide the initial binding site for preprotein docking (Allen, *et al.*, 2002, de Leeuw, *et al.*, 2002, Schnell and Hebert, 2003). It has also been proposed that the signal sequence is recognized by

TatC and then transferred to TatB (Alami, *et al.*, 2003). When signal peptide binding occurs, the PMF promotes the association between the TatBC complex and TatA oligomers. The folded preprotein is then translocated by the TatA channel and the leader peptide is processed. It has been shown that transport via the TAT pathway is less efficient (DeLisa, *et al.*, 2004) and slower than the Sec pathway with transit half-times in the order of a few minutes (Santini, *et al.*, 1998) instead of a few seconds (Berks, *et al.*, 2000) and the secretion machinery is rapidly saturated (Barrett, *et al.*, 2003, DeLisa, *et al.*, 2004).

2.2.6 Protein folding in the periplasm

Once proteins reach the periplasm, they encounter periplasmic chaperones, proteases, disulfide bond oxidoreductases/isomerases and peptidyl-prolyl isomerases (Fig. 3). Disulfide bond formation results from electron transfer pathways between the Dsb oxidoreductases (Kadokura, *et al.*, 2003). Depending on the function of the Dsb protein, they are kept either in the reduced (Dsb-$(SH)_2$; with a dithiol) or in the oxidized form (Dsb-S_2; with an intramolecular disulfide bond). The Dsb proteins are involved in two major pathways: an oxidation pathway (DsbA and DsbB) and an isomerization pathway (DsbC, DsbG, and DsbD).

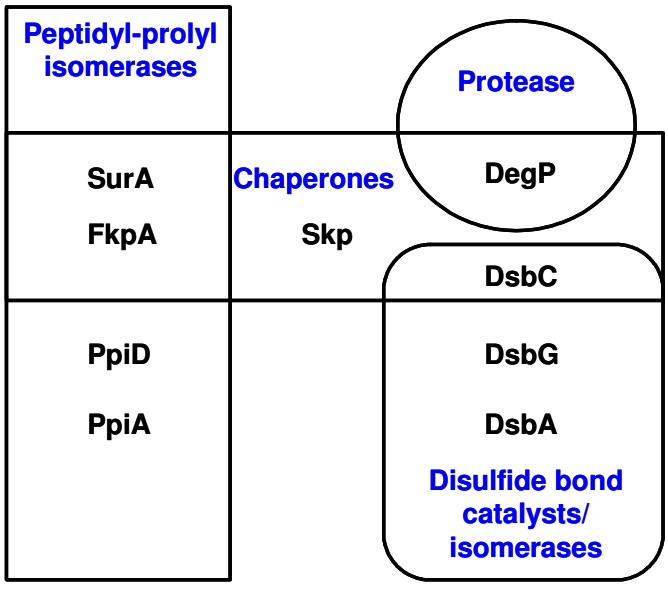


Fig. 3. Periplasmic folding catalysts have overlapping functions (Duguay and Silhavy, 2004)
The PPIases SurA and FkpA also have chaperone activity, as does the periplasmic protease, DegP, and the disulfide bond catalyst, DsbC. All of these proteins have chaperone activity. The reasons for this redundancy are unclear.

2.2.6.1 Disulfide bond formation

Until the early 1990s, disulfide bonds were thought to be formed in the periplasm spontaneously by molecular oxygen. However, the discovery of mutations in the *dsbA* gene revealed that disulfide bond formation is an enzyme catalyzed process (Bardwell*, et al.*, 1991, Kamitani*, et al.*, 1992)

DsbA: DsbA is a 21 kDa monomer containing a characteristic Cys30-Pro31-His32-Cys33 catalytic motif, embedded in a thioredoxin-like fold (Bardwell*, et al.*, 1991, Martin*, et al.*, 1993). DsbA is the strongest thiol oxidant with a standard redox potential of -119 mV (Zapun*, et al.*, 1993) which catalyses the oxidation of cysteines in the folding proteins. The high redox potential of DsbA arises from the unusual low pK_a value of about 3.5 of Cys30 (Nelson and Creighton, 1994). Thus it is entirely in a thiolate anion state under physiological conditions. The disulfide bond present in DsbA is very unstable and can be rapidly transferred to newly translocated reduced proteins (Guddat*, et al.*, 1998, Schirra*, et al.*, 1998). The flow diagram of the oxidation pathway in the periplasm is shown in Fig. 4. A bioinformatics search showed that more than 300 periplasmic proteins are potential substrates for DsbA (Hiniker and Bardwell, 2004).

DsbB: After transfer of its disulfide bond to a target protein, DsbA is released in the reduced form. To start a new catalytic oxidation cycle, DsbA must be reoxidized (Fig. 4). This reoxidation is accomplished by its partner enzyme, DsbB (Bardwell*, et al.*, 1993). DsbB is a 20 kDa inner-membrane protein that is predicted to have four transmembrane helices and two periplasmic loops (Jander*, et al.*, 1994). DsbB has four conserved cysteines, arranged in two pairs, which are required for activity. Both pairs are located in the periplasm on two different loops: Cys41 and Cys44 in the N-terminal loop and Cys104 and Cys130 in the C-terminal loop. After the reoxidation of DsbA, DsbB needs to pass the electrons it gained from DsbA to an acceptor. Under aerobic conditions, the ultimate electron acceptor is oxygen via quinone and cytochrome oxidase (Bader*, et al.*, 1999). Under anaerobic conditions, electrons are transferred from DsbB to menaquinone then to fumarate or nitrate reductase (Bader*, et al.*, 1999, Takahashi*, et al.*, 2004).

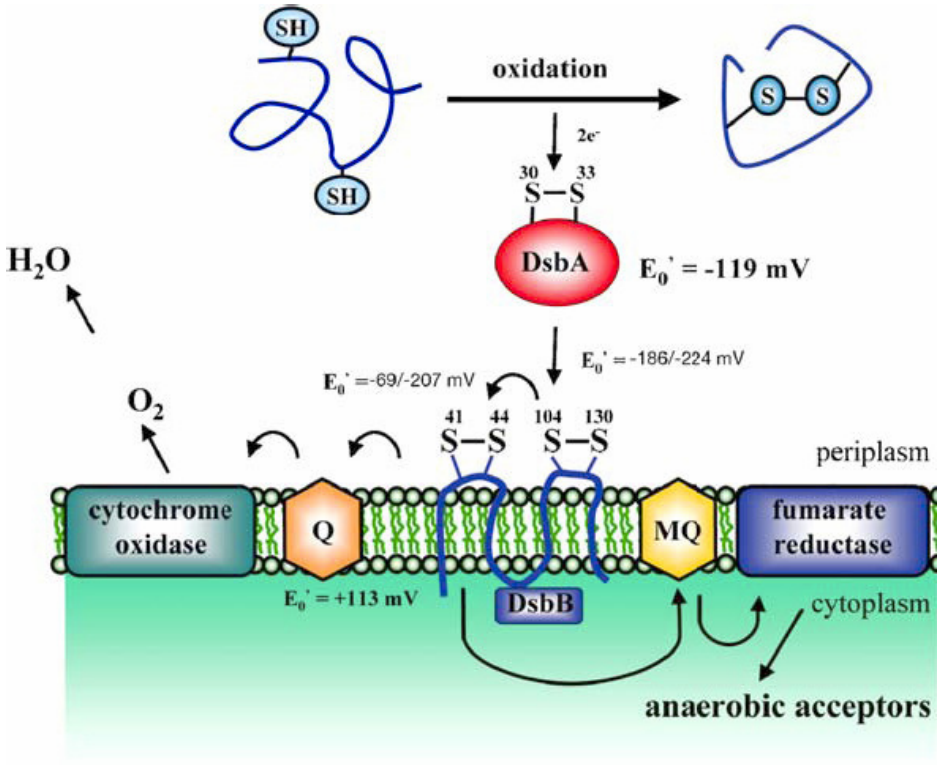


Fig. 4. The pathway of protein disulfide bond formation (Messens and Collet, 2006)
The direction of electron flow is shown by black arrows. DsbA reacts with a newly translocated protein. Free thiols in this protein are oxidized to form a disulfide bond. After donating its disulfide bond to a target protein, DsbA is released in the reduced form. To start a new catalytic oxidation cycle, DsbA must be reoxidized. This reoxidation is accomplished by its partner enzyme DsbB, a quinone reductase. Then, electrons flow from DsbB to ubiquinones and then to terminal oxidases, such as cytochrome b_d and b_o oxidases. The terminal oxidases transfer the electrons to oxygen in reactions coupled to H^+ transfer and production of H_2O. Under anaerobic conditions, DsbB passes electrons from DsbA onto menaquinone (MQ), which is up-regulated upon oxygen depletion. Anaerobic oxidoreductases such as fumarate reductase serve to reoxidize menaquinone.

2.2.6.2 Disulfide bond isomerization

DsbA is a powerful non-specific oxidant and quickly reacts with the cysteines of unfolded proteins as they enter in the periplasm. During this process, non-native disulfide bonds might be formed in the proteins that contain more than two cysteines. These incorrect disulfides have to be corrected to prevent protein misfolding. To correct non-native disulfides, *E. coli* possesses a disulfide isomerization system.

DsbC: DsbC is V-shaped homodimeric protein with two 23.3 kDa subunits. Each monomer consists of two domains: a C-terminal domain with a thioredoxin fold and a N-terminal dimerization domain (McCarthy, *et al.*, 2000). Each subunit contains four conserved cysteine

residues. Only two of them (Cys98 and Cys101), arranged in a CXXC motif, are essential for the isomerase activity. Cys98 has a low pK_a and is therefore in the thiolate form at neutral pH. Cys98 performs a nucleophilic attack on a non-native disulfide bond (Fig. 5). The reaction results in the formation of an unstable mixed disulfide complex between DsbC and the substrate. This mixed disulfide will be resolved either by attack of another cysteine of the misfolded protein, resulting in the formation of a more stable disulfide in the substrate and the release of reduced DsbC, or by the attack of the other cysteine of the CXXC motif, Cys101. In this latter case, DsbC is released in an oxidized state and will need to be reduced by DsbD. DsbC is particularly important for the folding of proteins with non-consecutive disulfide bonds (Berkmen*, et al.*, 2005). In addition to the isomerase activity, DsbC has chaperone activity *in vitro* (Chen*, et al.*, 1999).

DsbG: DsbG is a second protein disulfide isomerase present in the periplasm. It is about four-fold less abundant than DsbC. DsbG is a V-shaped homodimer (25.7 kDa subunit) and shares 24% amino acid identity with DsbC. DsbG has two conserved cysteine residues, Cys109 and Cys112, which are present in a CXXC motif (Bessette*, et al.*, 1999). DsbG has a redox potential similar to that of DsbC (Bessette*, et al.*, 1999, van Straaten*, et al.*, 1998) (Fig. 5). The size and surface charge of DsbG suggest that its substrates are larger and have less hydrophobic surfaces than the substrates of DsbC, which might indicate that DsbG preferentially catalyze disulfide bond isomerization in folded or partially folded proteins (Heras*, et al.*, 2004).

DsbD: DsbC and DsbG are kept reduced in the oxidizing environment of the periplasm by an inner-membrane protein called DsbD. DsbD consists of 546 amino acids and has three different domains: an N-terminal domain (α), a transmembrane domain (β) and a C-terminal domain (γ). Two of these domains α and γ, are in the periplasm, whereas the third one, β, has eight transmembrane segments. Each domain contains one pair of invariant cysteine residues (Fig. 5) that are essential for activity (Stewart*, et al.*, 1999). The function of DsbD is to give electrons to DsbC and DsbG which itself receives electrons from cytoplasmic thioredoxin. In the cytoplasm, thioredoxin kept reduced using electrons from NADPH (Rietsch*, et al.*, 1997).

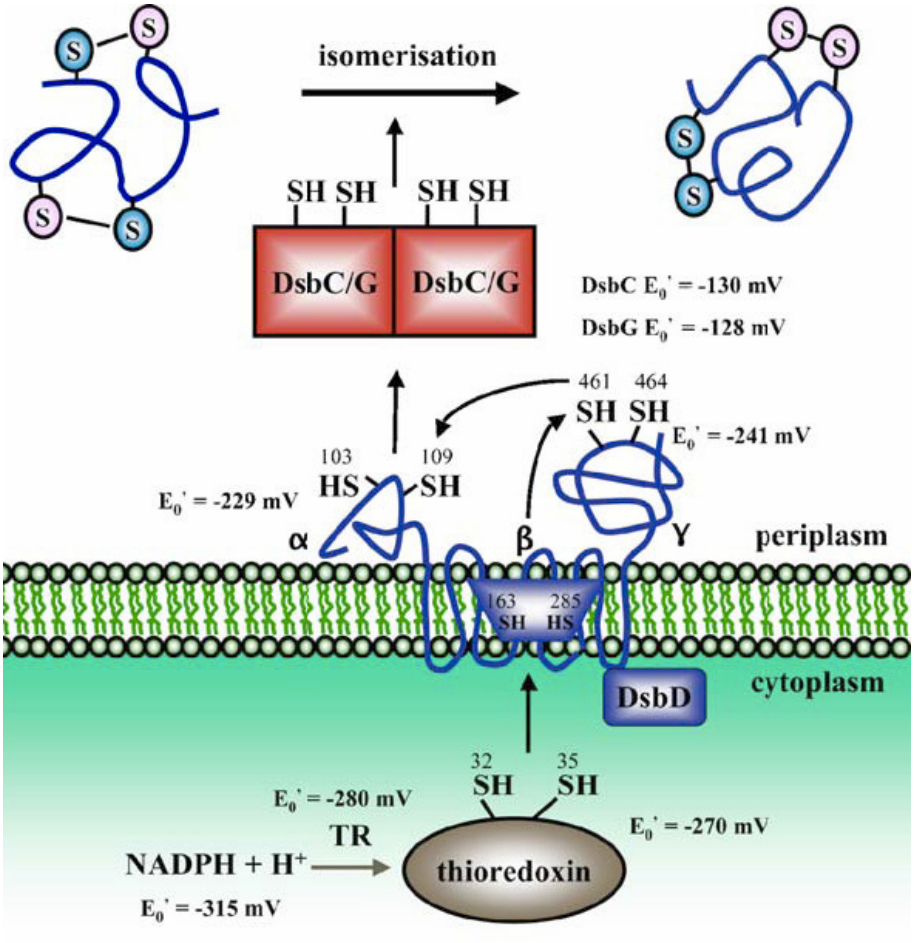


Fig. 5. The isomerization pathway (Messens and Collet, 2006)
The direction of electron flow is indicated by the arrows. Disulfide-bond rearrangement is catalyzed by the thiol–disulfide oxidoreductases DsbC and DsbG, which are maintained in a reduced state by the membrane protein DsbD. DsbD is reduced by cytoplasmic thioredoxin, which is recycled by thioredoxin reductase (TR) in a NADPH-dependent manner. In DsbD, the electrons flow from the membranous (β domain) to the C-terminal domain (γ domain) and then to the N-terminal domain (α domain).

2.2.6.3 Peptidyl-prolyl cis/trans isomerases

Peptidyl-prolyl bonds within proteins can exist either as cis- (~10%) or trans-isomers (Gothel and Marahiel, 1999). In the unfolded state, these bonds exist as a mixture of both isomers, but proper folding requires that all of the peptidyl-prolyl bonds are in the proper isomeric form which is characteristic for a specific protein. In the absence of catalysts, cis-trans isomerization is a slow process with activation energy of about 20 kcal/mol that is thought to be the rate-limiting step in protein folding (Brandts, *et al.*, 1975). This step is catalyzed by peptidyl-prolyl isomerases that are widely conserved in nature (Lang, *et al.*, 1987). Thus far, nine different PPIases have been identified in *E. coli*. Of these, five are in the cytoplasm, and four are in the periplasm.

Representatives of the three unrelated families of PPIases have been identified in the periplasm: PpiA also known as CypA or RotA is a cyclophilin that is not inhibited by cyclosporin A (Liu and Walsh, 1990). FkpA is related to the FK506-binding protein family (Horne and Young, 1995). PpiD (Dartigalongue and Raina, 1998) and SurA (Rouviere and Gross, 1996) are two parvulin homologs. A significant phenotype for null mutants has been found only for SurA indicating that the periplasmic PPIases are not essential for viability or have overlapping functions (Arie*, et al.*, 2001).

2.2.6.4 The periplasmic chaperones

Few periplasmic chaperones have been identified, and there are no classical Hsp chaperones. Indeed, as the periplasm lacks ATP, periplasmic chaperones must be mechanistically distinct from their cytoplasmic counterparts, most of which use ATP to drive their cycles of substrate binding and release. The chaperone activity of DegP was demonstrated *in vivo* and *in vitro* (Spiess*, et al.*, 1999). Genetic selections based on σE activity have identified the *surA* and *skp*/*ompH* genes (Missiakas*, et al.*, 1996, Rouviere and Gross, 1996). They are involved in the folding and assembly of outer membrane proteins (Schafer*, et al.*, 1999). *In vitro* chaperone activity of several periplasmic proteins such as DsbC (Chen*, et al.*, 1999) and DsbG (Shao*, et al.*, 2000) or substrate-binding proteins (e.g. OppA and MBP of *E. coli*) (Richarme and Caldas, 1997) have already been reported.

2.2.6.5 Proteolysis

One fate for misfolded periplasmic proteins is degradation by proteases. Although more than 10 periplasmic proteases have been identified in *E. coli*, DegP also known as HtrA or Do is the only protease identified as a heat shock protein involved in the degradation of misfolded proteins (Lipinska*, et al.*, 1990, Strauch*, et al.*, 1989). It has been shown that the activity of DegP could switch between chaperone and protease activities in a temperature-dependent manner (Spiess*, et al.*, 1999). At temperatures below 28 °C, DegP acts as a chaperone, protecting misfolded proteins from irreversible aggregation, and above 28 °C, its protease activity dramatically increases the degradation of misfolded proteins.

2.3 Periplasmic fusion proteins

In recent years it has become clear that fusion tags can have a positive impact on the yield and solubility of their fusion partners. In general, it is difficult to decide the best fusion system for a specific protein of interest. This depends upon the target protein itself (e.g. size, stability, and hydrophobicity), the expression compartment, and the application of the purified protein. The increase in yield and solubility for a given protein after fusion to a second protein or tag is different for each fusion. The exact mechanism by which fusion proteins enhance expression is not well understood. The hypothesis includes: (A) Fusion of a stable or conserved structure to an insoluble recombinant protein may serve to stabilize and promote proper folding of the recombinant protein (Butt*, et al.*, 2005) (B) Fusion tags may act as a nucleus of folding "molten globule hypothesis" (Creighton, 1997a). An efficient periplasmic fusion system should have the following features: (i) efficient translocator; (ii) enhance expression, solubility and proper folding; (iii) one-step purification; (iv) online or quick quantification; (v) reduce proteolysis; (vi) a minimal effect on the tertiary structure and biological activity; (vii) easy and specific removal of the fusion tag; (viii) applicability to a number of different proteins. However, no single fusion tag is optimal with respect to all of these parameters. Available periplasmic fusion proteins have their merits and demerits which are discussed in the following sections.

2.3.1 Maltose-binding protein (MBP)

MBP is a relatively large (40.6 kDa) protein localized in the periplasm and devoid of cysteine residue (Duplay*, et al.*, 1984). MBP is known for its remarkable solubility enhancement when it is fused at the N-terminus of model proteins (Sachdev and Chirgwin, 1998). Generally it is used for cytosolic expression but due to its natural localization in the periplasm, it is also used as periplasmic fusion tag for secretion, solubility and purification (Planson*, et al.*, 2003). It was observed that MBP remains soluble and properly folded when it is fused at the N-terminus of certain model proteins, but the passenger proteins are not properly folded and exist as in the state of soluble aggregates (Nallamsetty*, et al.*, 2005, Nomine*, et al.*, 2001, Sachdev and Chirgwin, 1999). The affinity of MBP for maltose is ~1 µM (Betton and Hofnung, 1996). Thermodynamically MBP is moderately stable with the T_m of 62.8 °C at pH 8.3 (Novokhatny and

Ingham, 1997)) and individual components of MBP fusions are slightly more stable than their counterparts in the fusion protein (Blondel*, et al.*, 1996).

2.3.2 Staphylococcal protein A (SpA)

SpA is a 31 kDa protein present on the surface of *Staphylococcus aureus*, that binds strongly and specifically to the constant part (F_c) of Immunoglobulins (Ig) (Cedergren*, et al.*, 1993). SpA does not contain any cysteine residues. SpA is highly soluble and renatures efficiently. Several smaller variants of SpA have been created. An engineered B domain (called as Z-domain) of SpA was described to be optimum as a fusion tag (Nilsson*, et al.*, 1987). A number of different expression vectors with or without signal sequences of SpA for the production of Z (7 kDa) or ZZ (14 kDa) fusions in which the fusion protein is either kept intracellularly (Nilsson*, et al.*, 1996, Nilsson*, et al.*, 1994) or exported to the periplasm or to the medium (Hammarberg*, et al.*, 1989, Stahl*, et al.*, 1989) have been developed. The small variants of SpA have advantages as a fusion partner over the full length SpA proteins (Nilsson*, et al.*, 1997). First, efficient secretion to the *E. coli* culture medium is obtained. Second, small variants bind to IgG with lower affinity than SpA. Furthermore, the small variants are considerably smaller than SpA and avoid steric hindrance in the affinity column.

2.3.3 Streptococcal protein G (SpG)

Streptococcal protein G (SpG) is a bifunctional receptor present on the surface of certain strains of streptococci and is capable of binding to both IgG and serum albumin (Nygren*, et al.*, 1988). The regions responsible for the affinities to serum albumin and IgG, respectively, are structurally separated. The serum albumin binding region of SpG contains three binding motifs (~5 kDa each), called ABD (albumin-binding domain). ABD (5-25 kDa) along with signal sequence of SpG has been utilized as fusion partners for intracellular as well as secreted proteins and can be purified in one-step via HSA-affinity chromatography. (Hammarberg*, et al.*, 1989, Larsson*, et al.*, 1996, Stahl*, et al.*, 1989).

2.3.4 Cellulose binding domain (CBD)

Cellulose binding domains of two enzymes (exoglucanase; nearly 100 residues and endoglucanase; 111 residues) from *Cellulomonas fimi* (Gilkes, *et al.*, 1988, Warren, *et al.*, 1986) were used as the fusion protein for the secretory production in *E. coli* (Creagh, *et al.*, 1996, Gilkes, *et al.*, 1992, Ong, *et al.*, 1991). CBD fusion system provides a very inexpensive ligand matrix (cellulose) for purification (Greenwood, *et al.*, 1989, Greenwood, *et al.*, 1992, Ong, *et al.*, 1991).

2.3.5 Disulfide bond oxidoreductase (DsbA)

In biotechnological applications, DsbA was used as a fusion tag assuming that DsbA should stabilize a fused unfolded polypeptide via its polypeptide binding site and promote correct disulfide bond formation by intra- and intermolecular catalysis. Bovine enterokinase catalytic subunit (Collinsracie, *et al.*, 1995) and human proinsulin (Winter, *et al.*, 2000) were produced in the native state in the periplasm after fusion with DsbA. Co-expression of DsbA only yielded periplasmic IGF-I inclusion bodies (Joly, *et al.*, 1998). The wild-type DsbA is an extremely efficient protein thiol oxidase and therefore causes the formation of aberrant disulfide bonds in proteins such as BPTI and hirudin (three disulfides) *in vitro* (Wunderlich, *et al.*, 1993, Zapun and Creighton, 1994).

2.3.6 Barnase

It is a relatively small (12.3 kDa), monomeric, extracellular RNAse from *Bacillus amyloliquefaciens* , without cysteines. An enzymatically inactive variant (H102A) of the RNAse barnase was used as a fusion protein for the translocation of cysteine-knot peptides to the periplasm. The majority of the peptides were in the native state (Schmoldt, *et al.*, 2005). Fusion of barnase to the C-terminus of antibody V_L domains, enhanced the solubility in the cytosol (Martsev, *et al.*, 2004). A simple one step purification of barnase fusion is possible via affinity chromatography on immobilized barstar, provided that the barnase H102A active site variant retains sufficient affinity to bind the inhibitor (Schmoldt, *et al.*, 2005).

2.4 Strategies for optimal expression

2.4.1 Molecular level

A number of central elements are essential in the design of recombinant expression systems (Baneyx, 1999, Jonasson*, et al.*, 2002). The genetic elements of the expression plasmid include origin of replication (*ori*), an antibiotic resistance marker, promoters, translation initiation regions (TIRs) as well as transcriptional and translational terminators, as shown in Fig. 6.

Most plasmid vectors used in recombinant protein expression replicate by the ColE1 or the p15A replicon. Plasmid copy number is controlled by the origin of replication (Baneyx, 1999). Different replicon incompatibility groups and drug resistance markers are required when multiple plasmids are employed for the coexpression of gene products. The most common drug resistance markers in recombinant expression plasmids confer resistance to ampicillin, kanamycin, chloramphenicol or tetracycline. Derivatives containing ColE1 and p15A replicons are often combined in this context since they are compatible plasmids (Mayer, 1995). The bacteriophage T7 polymerase based pET expression system is a very powerful tool for the expression of recombinant proteins in *E. coli* (Studier and Moffatt, 1986). The hybrid promoter *tac* is composed of the consensus -35 and -10 sequence from *trp* and *lac*UV5 promoters, respectively. The *trc* promoter is quite similar to *tac* promoter (Brosius*, et al.*, 1985) and used for high level of protein expression. A weak but tightly controlled expression system based on *araBAD* promoter that leads to a dose-dependent protein expression (Guzman*, et al.*, 1995) is often used for a more controlled production of proteins.

Codons that are rare in *E. coli* are often abundant in heterologous genes from distantly related sources (Kane, 1995). Expression of genes containing rare codons can lead to translational errors (McNulty*, et al.*, 2003). Site-directed mutagenesis of the target sequence (Kane*, et al.*, 1992) or co-transformation of a plasmid harboring a gene encoding the tRNA cognate to the problematic codons (Dieci*, et al.*, 2000) are potential solutions for the codon bias.

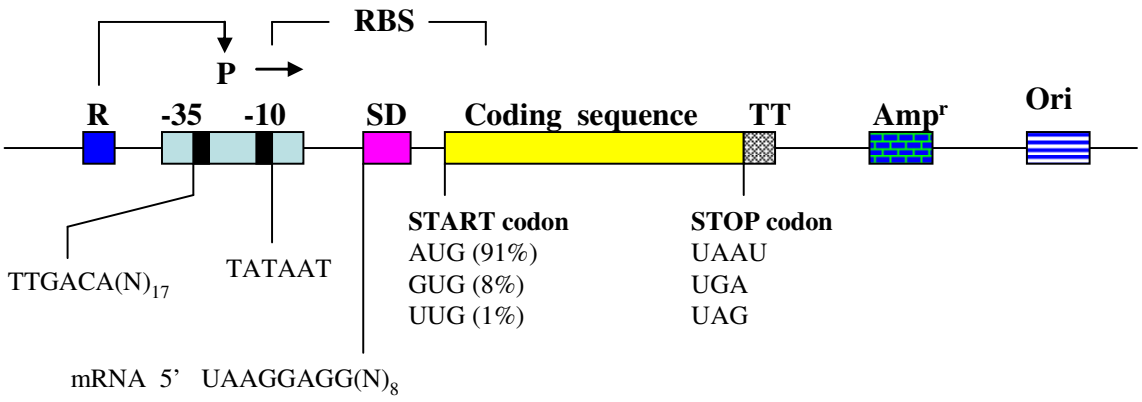


Fig. 6. Schematic presentation of the salient features and sequence elements of a prokaryotic expression vector (Makrides, 1996; modified)
Example is the *trc* promoter (P) consisting of the -35 and -10 sequences, which are separated by a 17-base spacer. The arrow indicates the direction of transcription. The RBS consists of the SD sequence followed by an A+T-rich translational spacer that has an optimal length of approximately 8 bases. The SD sequence interacts with the 3' end of the 16S rRNA during translational initiation, as shown. The three start codons are shown, along with the frequency of their usage in *E. coli*. Among the three stop codons, UAA followed by U is the most efficient translational termination sequence in *E. coli*. The repressor is encoded by a regulatory gene (R), which may be present on the vector itself or may be integrated in the host chromosome, and it modulates the activity of the promoter. The transcription terminator (TT) serves to stabilize the mRNA and the vector. In addition, an antibiotic resistance gene, e.g., ampicillin, facilitates phenotypic selection of the vector, and the origin of replication (Ori) determines the vector copy number.

2.4.2 *E. coli* host strains

The strain or genetic background for recombinant expression is highly important. Expression strains should be deficient in the most harmful natural proteases, maintain the expression plasmid stably and confer the genetic elements relevant to the expression system (e.g., DE3). Advantageous strains for a number of individual applications are available. *E. coli* BL21 is the most common host for research purposes. BL21(DE3) is a robust *E. coli* B strain, able to grow vigorously in minimal medium. BL21 is deficient in ompT and lon proteases, thus at least some target proteins should be more stable than in host strains containing these proteases.

2.4.3 Stress response

Stress responses during expression of proteins may be caused by environmental stress situations such as heat shock, amino acid depletion or starvation (Bailey, 1993). Some recombinant proteins directly influence the host cellular metabolism by their enzymatic properties, but in general expression of recombinant proteins induces a "metabolic burden" (Bentley and Kompala, 1990).

Stress can be reduced in recombinant systems by slow adaptation of cells to a specific production task. This can be accomplished by gradually increasing the level of inducer or by slowly increasing the plasmid copy number during cultivation (Trepod and Mott, 2002).

2.4.4 Fermentation optimization

Protein production in *E. coli* can be increased significantly by the use of high cell density culture system, in which cell concentrations can achieve in excess of 100 g (dry cell weight)/liter and can provide a cost-effective production of recombinant proteins. Detailed reviews of large-scale fermentation systems have been published (Choi*, et al.*, 2006, Lee, 1996, Thiry and Cingolani, 2002). The composition of the cell growth medium must be carefully formulated and monitored, because it may have significant metabolic effects on both the cells metabolism and protein production. Nutrient composition and fermentation variables such as temperature, pH and other parameters can affect proteolytic activity, secretion, and production levels (Baneyx and Georgiou, 1992). Addition of low molecular weight folding enhancers in the medium can cause more than a 400-fold increase in the production of soluble, active protein (Blackwell and Horgan, 1991).

High cell density culture systems suffer from several drawbacks, including limited availability of dissolved oxygen at high cell density, carbon dioxide levels which can decrease growth rates and stimulate acetate formation, reduction in the mixing efficiency of the fermentor, and heat generation. The techniques that are used to minimize such problems have been examined in detail (Lee, 1996). A major challenge in the production of recombinant proteins at high cell density is the accumulation of acetate that is detrimental to cell growth (Jensen and Carlsen, 1990, Lee, 1996, Luli and Strohl, 1990). A number of strategies have been developed to reduce acetate formation in high cell density cultures, but these suffer from several drawbacks (Lee, 1996). The promising solution to solve the acetate accumulation is incorporation of *alsS* gene from *B. subtilis* encoding acetolactate synthetase which catalyses the conversion of pyruvate to nonacidic and less toxic byproducts into *E. coli* cells (Aristidou*, et al.*, 1995, San*, et al.*, 1994). In this strain of *E. coli*, level of acetate accumulation was reduced while production of recombinant protein was significantly improved (Aristidou*, et al.*, 1995, San*, et al.*, 1994).

3 Objective

Efficient translocation and proper folding are two major intrinsic limitations for the high level periplasmic production of recombinant proteins. The aim of the present work was to overcome the bottleneck of translocation and to enhance *in vivo* folding in the periplasm. These are important aspects to obtain a maximal yield. Two model proteins, human pepsinogen and human proinsulin, were used. Both proteins differ in size, folds and function but both contain multiple disulfide bonds.

In the present work, human pepsinogen and human proinsulin were directed to *E. coli* periplasm using different approaches: (i) use of promoters of various strengths; (ii) addition of different signal sequences to follow either co- or post-translational translocation; (iii) use of different *E. coli* strains; (iv) choosing a variety of cultivation conditions; (v) addition of low molecular weight additives; (vi) and by exploring a new fusion protein system.

4 Results and Discussion

The data obtained for the periplasmic production of disulfide bonded proteins, human pepsinogen and human proinsulin, are summarized in the following sections. Details of the data are published and presented in the chapter 7.

4.1 Development of a fluorogenic assay for pepsin

Pepsinogen (40.1 kDa) is a zymogen of pepsin (34.6 kDa) containing three consecutive disulfide bonds required for proper folding and activity. Acidification of pepsinogen leads to its autoactivation, resulting in the formation of pepsin. Enzymatic activity was performed for the quantification of properly folded pepsinogen in the periplasmic extract due to close structure-function relationship.

Usually, the expression level of properly folded proteins in the periplasm is in the order of ng/ml. Therefore, a very sensitive assay was desired for the precise quantification of pepsinogen in the periplasmic extracts. Several existing pepsin assays were evaluated using commercial porcine pepsinogen as a reference.

4.1.1 Evaluation of various pepsin assays

Pepsin is a well known aspartic protease. It was partially purified, characterized and crystallized in 1920's (Northrop, 1930). With time, different assays were developed according to need. The substrates for pepsin are based on use of hydrophobic amino acids in the peptides (Inouye and Fruton, 1967), labeled with chromo- or fluorophore, or whole proteins e.g. hemoglobin, casein etc. (Anson, 1938) or fluorophore-quenched proteins (Jones*, et al.*, 1997). The most sensitive assay for pepsin is the use of radiolabelled hemoglobin (Lin*, et al.*, 1989) but this assay is not convenient for routine purpose due to safety reasons and requirement of special instruments. In the present study, four different pepsin assays were tested. The results are described and discussed below.

Use of a chromogenic peptide substrate: Pepsin is a broad range protease but preferentially hydrolyzes peptide bonds between large hydrophobic aromatic residues (Inouye and Fruton, 1967). Therefore, all the peptides used in pepsin assay are hydrophobic although hydrophilic groups are introduced to increase the solubility of the peptides. These peptides still have significant hydrophobicity and cause solubility problems. In this study, Phe-Ala-Ala-4-nitro-Phe-Phe-Val-Leu-pyridin-4-ylmethylester (cat. no: M-1690 from BACHEM) was used as substrate. Pepsin cleaves in between two phenylalanine residues and liberates nitro phenyl group at the end. A yellow colored solution develops. Using 100 µM substrate and 40 nM pepsin in 40 mM formate buffer, pH 3.5 at room temperature, the total increase in absorbance at 310 nm was only 0.05. Due to the low sensitivity and solubility, other assays were examined. These were based on whole protein as substrate.

QuantiCleave™ protease assay kit from Pierce: In this assay, the primary amines on the surface of casein are blocked with succinic anhydride. In the presence of proteases, succinylated casein is cleaved thereby liberating primary amines at the end of peptides. TNBSA at alkaline pH reacts with exposed primary amines to produce an orange-yellow color with absorption at 450 nm. Since succinylated casein was precipitating below pH 4.0, digestion of succinylated casein with pepsin was performed at pH 4.5. The quantification range of porcine pepsin was 100-500 ng. However, it lead to only a little increase in absorption (ΔA_{450} = 0.05). Moreover, this assay was not suitable for quantification of pepsinogen in the periplasmic extracts due to interference of primary amines from the buffer (e.g. Tris) and due to the high protein content.

EnzChek® Protease assay kit from Molecular Probes: In this assay, casein is heavily labeled with the pH-insensitive fluorescent BODIPY® FL dye, resulting in an almost quantitative quenching of the conjugate's fluorescence. After digestion with pepsin at pH 2.0, liberated peptides with conjugated BODIPY® FL exhibited fluorescence with an excitation/emission maxima at 505/513 nm. This assay was sensitive for the quantification of pepsin up to 10 ng in the microtiter plate format. However, it was found that the initial increase in fluorescence was independent from the pepsin concentration. Furthermore, the presence of *E. coli* proteases in the periplasmic extracts of negative control showed significant background.

Use of hemoglobin: Hemoglobin is a commonly used substrate for pepsin quantification (Anson, 1938). A 2% hemoglobin solution in 60 mM HCl was incubated with pepsin at 37 °C. Uncleaved hemoglobin was precipitated by addition of 5% TCA. The absorbance of soluble peptides was measured at 280 nm. The sensitivity of this hemoglobin based pepsin assay was low (1-10 µg/ml). It was observed that the presence of nucleic acids in the protein sample masked the absorption of TCA soluble peptides at 280 nm. Refolded probes from inclusion bodies or periplasmic extracts often contain nucleic acids. Therefore, the hemoglobin based pepsin assay was not suitable.

The milk clotting assay for pepsin quantification is a sensitive assay in the ng range. However, the turbidity of milk can be changed by a different protease and other agents (McPhie, 1976). Therefore, this assay is not precise and selective for pepsin. Intramolecular fluorophore quenched hydrophobic peptides in the pepsin assay are also sensitive in the ng range (da Silva Gomes, *et al.*, 2003), but due to low the solubility and high costs of these fluorogenic peptides they were not evaluated for routine estimation of pepsinogen from periplasmic extracts.

4.1.2 Use of EGFP as a substrate

In the next step, a well known and characterized fluorogenic protein (GFP) was tested as a potential substrate for pepsin quantification. EGFP is a variant of GFP with 35-fold more fluorescence intensity in comparison to the wild-type protein (Cormack, *et al.*, 1996). EGFP is itself very soluble and stable. Nearly the entire protein is required for fluorophore formation (Li, *et al.*, 1997). Therefore, digested EGFP does not exhibit fluorescence. GFP has a strong tendency to renaturate (80-90%) in the absence of denaturants (Ward and Bokman, 1982). Based on these informations, a new assay for pepsin was developed.

4.1.3 Reversible unfolding of EGFP

The optimum pH of pepsin activity is 1.0 for substrates like casein and hemoglobin (Cornis-Bowden. A. J. and Knowles, 1969). Therefore, EGFP must be acidified in order to be cleaved by pepsin. EGFP exhibits stable fluorescence at alkaline pH and loses its fluorescence at acidic pH

due to denaturation. Since EGFP retains only 50% fluorescence at pH 5.9 (Patterson*, et al.*, 2001). Therefore, remaining intact EGFP molecules must be neutralized after pepsin cleavage.

GFP exhibits a pH-independent fluorescence in the range of pH 7-10 (Bokman and Ward, 1981). The dependence of EGFP fluorescence on its concentration was determined. The fluorescence of native EGFP was linearly dependent on the protein concentration up to 55 µg/ml (Fig. 1A, chapter 7.1). To analyze the appropriate pH for acid denaturation, incubation time and temperature, 11 µg of EGFP was acidified to pH 2.0, 3.0, 3.5 and 4.0 by adding 0.1 volumes of the 1 M respective citrate buffer at five different temperatures (20, 30, 37, 45 and 55 °C). After 10, 30 and 60 min of denaturation, EGFP was renatured by adding 0.4 M Tris-HCl, pH 8.5. As shown in Fig. 7, EGFP underwent complete renaturation when it was denatured to pH 2.0 at 20-37 °C up to 60 min. When EGFP was denatured above pH 3.0, regaining of fluorescence was reduced at all the temperatures and incubation times.

To analyze the concentration dependent EGFP renaturation, varying concentrations of EGFP up to 55 µg/ml were acidified to pH 2.0 for 10 min at room temperature and then neutralized again to pH 8.0. The fluorescence of EGFP obtained after renaturation was the same as that of the native EGFP, indicating the complete refolding under these conditions (Fig. 1A, chapter 7.1). After neutralization of acidified EGFP at room temperature, fluorescence was completely recovered within 5 min indicating fast refolding, and the fluorescence of refolded EGFP was stable for at least 1 hour (Fig. 1B, chapter 7.1)

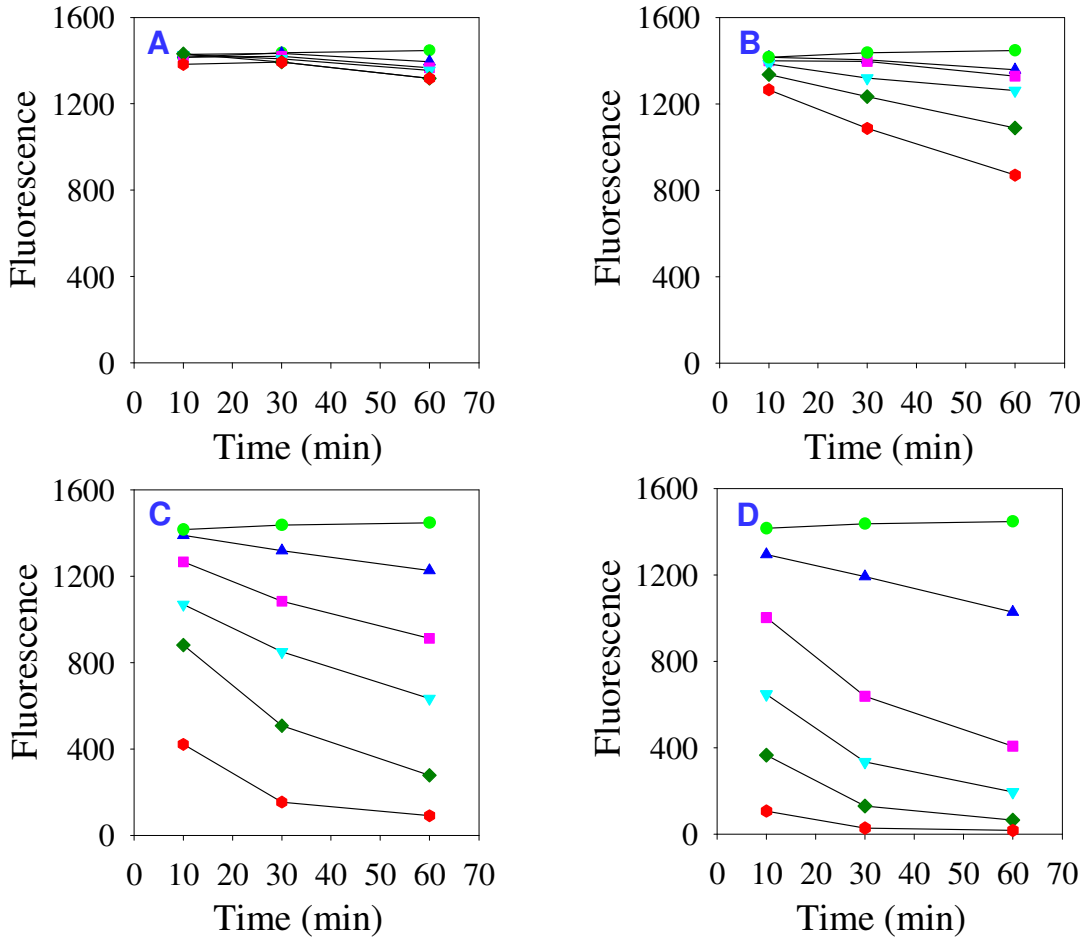


Fig. 7. Effect of denaturating pH, temperature and incubation time on the refolding of EGFP.
(A) 11 µg EGFP was acidified by adding 0.1 volume of 1 M citrate buffer pH 2.0 at 20 °C (▲), 30 °C (■), 37 °C (▼), 45 °C (◆) and 55 °C (●). After incubation for 10, 30 and 60 min, the solution was neutralized to pH 8.0 by adding 0.4 M Tris-HCl, pH 8.5. Finally, the fluorescence of each probe was measured. (●) represents the fluorescence of 11 µg native EGFP at pH 8.0. Excitation was done at 490 nm and emission was measured at 508 nm. Figures B, C and D represent the denaturation of EGFP at pH 3.0, 3.5 and 4.0, respectively and renaturation to pH 8.0, keeping the other parameters and symbols the same.

4.1.4 Sensitivity of the EGFP assay

To determine the sensitivity of the assay, 11 µg acidified EGFP at pH 2.0 was digested with varying concentrations of pepsin. After neutralization, undigested EGFP molecules were quantified by measuring its fluorescence. The decrease in the fluorescence of EGFP was proportional to the amount of added pepsin, incubation temperature and time. After digestion of EGFP at 37 °C for 3 hours followed by neutralization, 50 pg of pepsin were sufficient to significantly reduce amount of EGFP fluorescence (Fig. 2, chapter 7.1). When 5 µM pepstatin was added before addition of pepsin in the acidified EGFP solution, the fluorescence was

completely regained due to inhibition of pepsin activity. For the standard condition at pH 2.0, 10 min of EGFP digestion at 20 °C was sufficient to generate the standard curve for the quantification of 0-30 ng pepsin (Fig. 8).

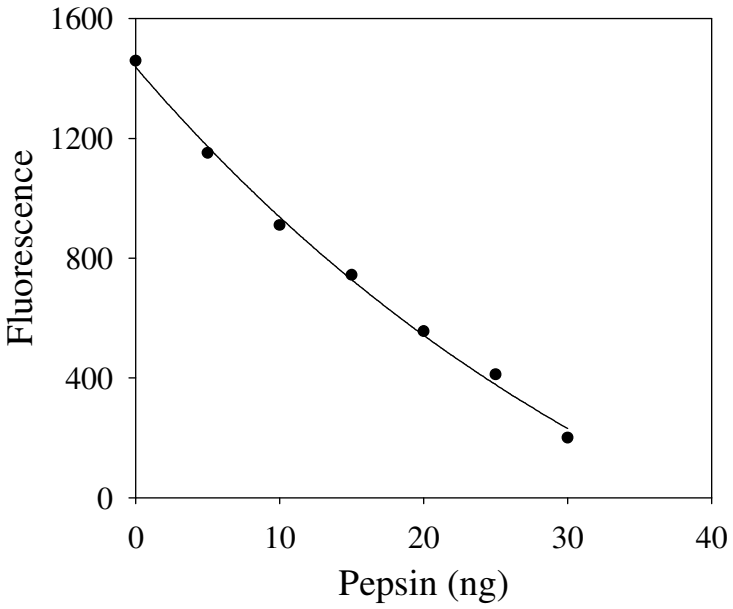


Fig. 8. Standard curve of EGFP proteolysis by pepsin.
11 µg of EGFP was acid denatured at pH 2.0. Next, 0-30 ng pepsin were added and the samples were incubated at 20 °C for 10 min. Remaining intact EGFP molecules were renatured by adding 0.4 M Tris-HCl, pH 8.5 and measurement of renatured EGFP was determined with excitation at 490 nm and emission at 508 nm.

4.1.5 Kinetics of proteolysis

The specific digestion rates of EGFP (55 µg/ml) was determined using three different concentrations of pepsin (4.8, 14.3 and 23.8 ng/ml) and activated pepsinogen (4.8, 14.4 and 24.0 ng/ml) in the time-dependent proteolysis. The specific rates of digestion for pepsin (38.2 ± 6.7 ng EGFP ng^{-1} pepsin min^{-1}) and activated pepsinogen (37.2 ± 5.2 ng EGFP ng^{-1} activated pepsinogen min^{-1}) were quite similar (Fig. 3 A-B, chapter 7.1). Substrate-dependent protease activity showed a typical Michaelis-Menten kinetics (Fig. 4 A-B, chapter 7.1). The apparent V_{max} and K_m values for pepsin and activated pepsinogen (Table 1, chapter 7.1) were calculated from the corresponding double-reciprocal plots (Fig. 4 A-B, chapter 7.1).

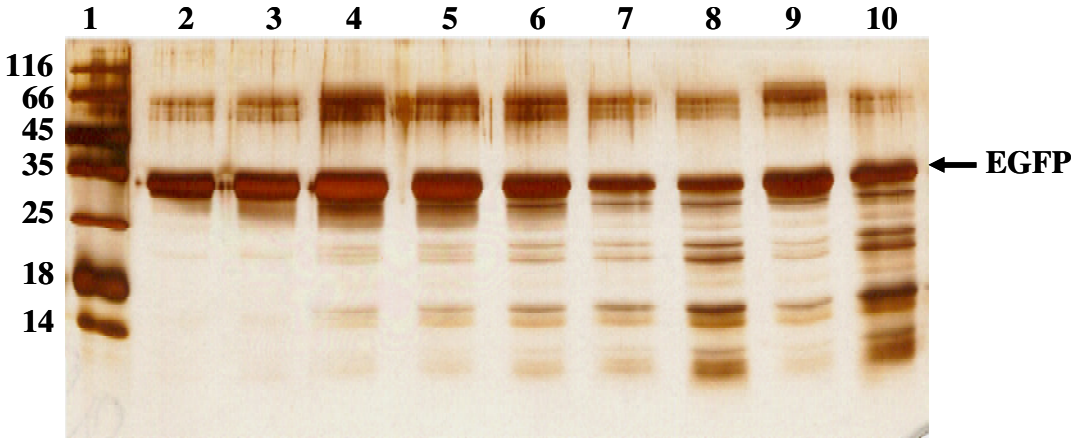


Fig. 9. SDS–PAGE analysis of pepsin-catalyzed EGFP cleavage.
Acid denatured EGFP (55 µg) was incubated with 24.0 ng pepsin or activated pepsinogen. At different time intervals, proteolysis was stopped by neutralization. The probes were subjected to TCA precipitation, and an aliquot (1.1 µg) was separated on 17.5% SDS–PAGE. Lane 1, standard proteins with the masses given in kDa; lane 2, 1.1 µg native EGFP; lanes 3–8, EGFP after digestion by pepsin at 20 °C for 0, 2, 6, 10, 20, and 30 min; lanes 9 and 10, EGFP after digestion by activated pepsinogen at 20 °C for 10 and 30 min.

The kinetic of EGFP proteolysis by pepsin and activated pepsinogen was visualized by SDS-PAGE. At different intervals of EGFP proteolysis by pepsin or activated pepsinogen, probes were precipitated by TCA and about 1.1 µg digested EGFP was loaded on to a 17.5% SDS-PAGE. As shown in Fig. 9, the same pattern of proteolysis was observed for both pepsin (lanes 3-8) or activated pepsinogen (lanes 9-10). EGFP contains more than 45 susceptible pepsin cleavage sites (http://www.expasy.ch/tools/peptidecutter). The appearance of some prominent proteolytic fragments of EGFP (Fig. 9) indicates the presence of certain preferential cleavage sites at pH 2.0. Nearly the entire protein sequence is required to produce fluorescence (Li*, et al.*, 1997). Thus, a single pepsin cut should be sufficient to inhibit refolding and restoration of the fluorophore.

Usually aspartic proteases have pH optima in the acidic range (Gruninger-Leitch*, et al.*, 2002); therefore acid denatured EGFP would be a good substrate for all the proteases working below pH 3.0. Due to a very compact structure, GFP in the folded state is very resistant to serine and metalloproteases at neutral pH, even at 1 mg/ml concentration (Bokman and Ward, 1981). Thus, after neutralization of EGFP, interference by common proteases present in the crude sample can not occur. In this assay, measurement is based on emission of fluorescence at 508 nm which omits the possibility of interference of nucleic acids present in the samples. EGFP can be recombinantly produced as soluble protein, easily be purified and can be stored at low temperature for long periods. The fluorophore of GFP is so tightly packed near the geometric

center of beta-can that classical fluorescence quenching agents such as acrylamide, halides and molecular oxygen have almost no effect on GFP fluorescence (Rao*, et al.*, 1980). GFP is stable with a relatively high T_m of 76 °C (Bokman and Ward, 1981, Ward and Bokman, 1982).

All the proteases of *E. coli* belong to the serine and metalloprotease family (Gottesman, 1996). Therefore, the EGFP based pepsin assay developed here is selective for the quantification of recombinant pepsinogen in the periplasmic extracts. This assay was routinely used for the precise estimation of pepsinogen in the periplasmic fractions of *E. coli*.

4.2 Periplasmic expression of human pepsinogen

4.2.1 Cloning strategy for periplasmic expression of pepsinogen

To produce pepsinogen in the periplasm of *E. coli*, various signal sequences from prokaryotic origin were chosen. pET12a and pET20b vectors (Novagen) contain the *ompT* and *pelB* signal sequence, respectively, upstream of the multiple cloning sites for periplasmic accumulation of cloned proteins. Human pepsinogen A was amplified from pHQPEX-30-5 (kindly provided by Dr. R. Bolli, ZLB Bioplasma AG, Switzerland) and unique restriction sites *Nco*I and *Sal*I were introduced at 5' and 3' ends for cloning in pET20b while *Sal*I site at both ends for cloning in pET12a vector. The blunt end PCR fragments were cloned into pCR-Blunt II-TOPO (Invitrogen). The clones were sequenced and digested with respective restriction enzymes and purified by agarose gel electrophoresis. The fragments were ligated into digested and dephosphorylated pET12a and pET20b vectors. The clones in pET12a and pET20b were sequenced and abbreviated as pET-ompT-peps and pET-pelB-peps, respectively. The signal sequence of DsbA was the third leader peptide for the translocation of pepsinogen to the periplasm. In contrast to other signal sequences, dsbA facilitates a co-translational export of passenger proteins to the periplasm via SRP pathway (Schierle*, et al.*, 2003). Therefore, it is suitable for translocation of fast folding passenger proteins to the periplasm. A synthetic nucleotide sequence encoding the signal sequence of *dsbA* was inserted into pET20b, thereby, replacing the signal sequence of *pelB* by *dsbA*. The digested pepsinogen fragment was ligated into pET20b-dsbA and the construct was named pET-dsbA-peps.

For the expression of pepsinogen under control of *trc* and *araBAD* promoters in pTrc99a and pBAD22 vectors, pepsinogen along with the *pelB* signal sequence was amplified from pET-pelB-peps. A hexa-histidine sequence at 3' end of pepsinogen was introduced by the PCR primer. The amplified fragment was cloned into pCR-Blunt II-TOPO and digested by *Eco*RI and *Sal*I. Subsequently, the pepsinogen fragment was purified by agarose gel electrophoresis and ligated into *Eco*RI-*Sal*I digested and dephosphorylated pTrc99a and pBAD22 vectors to give pTrc-pelB-peps and pBAD-pelB-peps, respectively.

In a new fusion approach for the translocation of pepsinogen to the periplasm of *E. coli*, the sequence of ecotin (GenBank M60876) was amplified from *E. coli* JM83. Three repeats of Gly-Ser sequence at 3' end of ecotin were introduced with a unique restriction site for *Bsp*E1 by the 3' primer. The fragment was cloned into pCR-Blunt II-TOPO. Pepsinogen was amplified from pHQPEX-30-5 and three repeats of Gly-Ser were introduced at 5' end while the hexa-histidine sequence was introduced at 3' end of pepsinogen. The fragment was cloned into pCR-Blunt II-TOPO. The ecotin fragment was cut with *Eco*RI-*Bsp*EI and pepsinogen fragment was cut with *Bsp*E1-*Sal*I. Both fragments were purified by agarose gel electrophoresis and ligated into *Eco*RI-*Sal*I digested and dephosphorylated pTrc99a and pBAD22 to give pTrc-eco-peps (abbreviated as pEGP1) and pBAD-eco-peps (abbreviated as pB-EGP1), respectively (Fig. 1B, chapter 7.2).

4.2.2 Expression of pelB-pepsinogen, ompT-pepsinogen and dsbA-pepsinogen

Experiments for the periplasmic production of pepsinogen were started directly after transformation and overnight incubation. *E. coli* BL21(DE3)Gold harboring different constructs (pET-pelB-peps, pET-ompT-peps or pET-dsbA-peps) was grown in 20 ml LB medium at 24 °C. Protein synthesis was induced with 1 mM IPTG at OD_{600} of 0.5-0.7. The cells were further incubated up to 20 hours at 24 °C. Cells harboring respective empty vector (pET12a, pET20b) were used as negative controls.

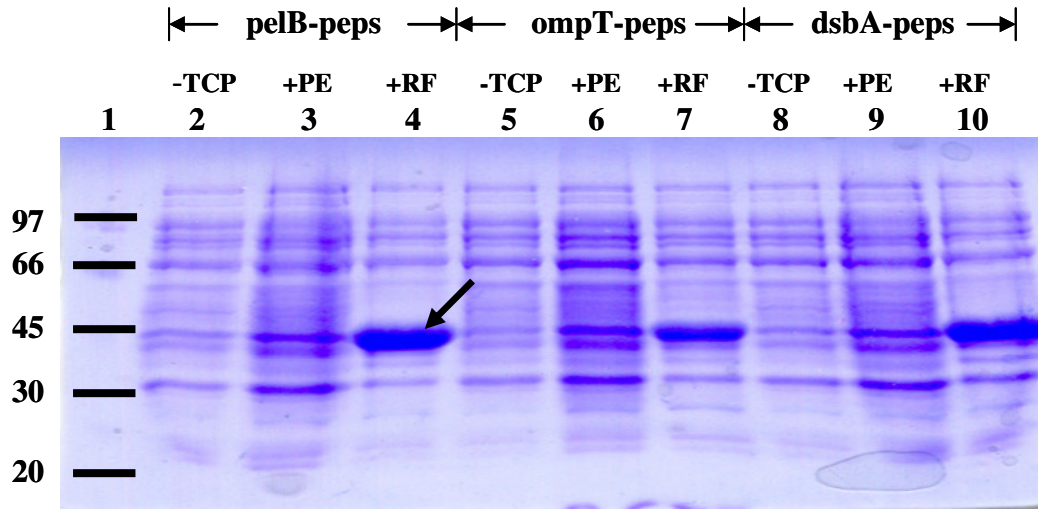


Fig. 10. Expression of pelB, ompT and dsbA-pepsinogen fusions.
Lane 1, protein marker in kDa, lane 2-4, Fraction of cells harboring pET-pelB-peps; lanes 5-7, fraction of cells harboring pET-ompT-peps and lanes 8-10 represents the fraction of cells harboring pET-dsbA-peps. (-) TCP, (+) PE and (+) RF indicates uninduced total cell protein, induced periplasmic extract and induced rest fractions of cells after periplasm extraction, respectively.

Expression of pepsinogen in the periplasm was analyzed 1, 3, 5 and 20 hours after addition of IPTG. The periplasmic extracts were prepared according to Winter *et al* (Winter*, et al.*, 2000). The periplasmic extract and the rest fraction of cells from these cultures were used for determination of pepsin activity and were analyzed by SDS-PAGE. As shown in Fig. 10, fusion of pepsinogen with pelB, ompT and dsbA were produced in high amounts but remained in the rest fraction. After separation of the probe by SDS-PAGE, they were blotted on to a PVDF membrane and the prominent band (marked with ↙, Fig. 10) was subjected to N-terminal sequencing. The band corresponded to the uncleaved pelB signal sequence. In the enzymatic assay, no pepsin activity was found in all the periplasmic extracts.

To avoid any misfolding or aggregation of translocated pepsinogen in the periplasm, several classes of folding enhancers were added in the culture medium and the inducer concentration (0.02-1 mM IPTG) was varied. After induction with 1 mM IPTG at mid-log phase, the following additives were added in the medium. Redox shuffling compounds (1-5 mM GSH, 1-5 mM GSSG and 0.01 - 0.1 mM vectrase P), solubilizing agents (0.4 M L-arginine, 5% ionic liquid salts†, 50 mM urea), stabilizing agents (0.1 M proline, 0.1 M sulfobetaine), or detergents (0.1% brij 58P, 0.1% triton X-100) were added individually to the cultures. In another set of experiments, after induction with 1 mM IPTG, the above mentioned additives were added keeping 1 mM GSH in all the cultures. Even after such manipulation, no pepsin activity could be determined.

The majority of periplasmic proteins are translocated via the SecB dependent pathway. In this pathway, proteins are post-translationally translocated and should remain in translocation-competent state. Hence, SecB dependent pathway is not efficient for fast folding proteins. DsbA signal sequence facilitates co-translational translocation so that fast folding proteins can be translocated via this pathway. However, pepsin activity could not be obtained in the periplasmic fraction even after using the dsbA signal sequence.

These data indicated two possible explanations. First, due to the strong T7 promoter, the translocation machinery was not able to support the high frequency of transcription and translation therefore pepsinogen along with signal sequences accumulated in the cytosol. Secondly, pepsinogen was translocated to the inner membrane but either due to improper activity

† [1-3 Dimethyl Imidazolium Methyl Sulphate and 1-(Hydroxy Ethyl) 3 Methyl Imidazolium Chloride]

of signal peptidase I or due to some peculiar property of pepsinogen, prevented it from crossing the negatively charged inner membrane barrier.

To exclude the problem associated with the strong T7 promoter, *E. coli* BL21(DE3)Gold harboring pTrc-pelB-peps and pBAD-pelB-peps plasmids were analyzed. Next, they were grown at 37 °C in 20 ml LB medium until mid-log phase. For the induction of protein synthesis, 1 mM IPTG was used for the *trc* promoter and 0.02% L-arabinose was used for the *araBAD* promoter. Cells harboring pTrc99a and pBAD22 vectors were used as a negative control. After induction with 1 mM IPTG and at 24 °C for 4 hours, nearly 16 µg pepsin per liter OD1 culture was obtained for cells harboring pTrc-pelB-peps. After addition of 5 µM pepstatin A, which is specific potent inhibitor of pepsin, the protease activity in the periplasmic fraction was abolished. No proteolytic activity was observed in the negative control. Different strains, BL21, JM109, C600 were used; varied inducer concentrations, 0.01-1 mM IPTG; media, LB and M9; pre-induction OD_{600} (0.2-1); and post-induction incubation times (3-5 hours) were employed but production of pepsinogen using the pelB signal sequence under control of the *trc* promoter was not reproducible. When pelB-pepsinogen was produced under control of the *araBAD* promoter, no significant pepsin activity was obtained in the periplasmic fraction probably due to a poor expression level.

4.2.3 Expression of pepsinogen upon fusion to *E. coli* ecotin

To analyze whether for the failure to obtain native pepsinogen in the periplasm, was associated with the processing of signal peptides or interference of the mature part of pepsinogen in the translocation. A periplasmic production of pepsinogen with the full-length ecotin was performed. For the expression of ecotin-pepsinogen, pEGP1 and pB-EGP1 plasmids were transformed in *E. coli* BL21(DE3)Gold. In the control experiment, *E. coli* BL21(DE3)Gold harboring pTrc99a and pBAD22 plasmids were grown and induced under identical conditions. To produce ecotin-pepsinogen under control of the *trc* promoter, 1 mM IPTG was used while the production under control of *araBAD* promoter was done in the presence of 0.02% L-arabinose at OD_{600} 0.5-0.8. The cells were incubated at 24 °C for 3.5 hours. Pepsin activity was analyzed in the periplasmic fractions using EGFP as a substrate (Chapter 7.2). The pepsin activity in the periplasmic fractions

of cells containing pEGP1 and pB-EGP1 plasmids corresponded to 100 and 8 µg ecotin-pepsinogen per liter OD1, respectively.

As stated earlier, even the protein itself is recognized by translocation machinery (Flower*, et al.*, 1994). The presence of basic residues in the first 18 residues of a mature protein severely affects the translocation across the cytoplasmic membrane in *E. coli* (Kajava*, et al.*, 2000). After fusion to ecotin this bottleneck of translocation could be overcome. After fusion to ecotin, pepsinogen was reproducibly produced in all the experiments. Due to the higher yield obtained with pEGP1, this plasmid was used for large scale expression, purification and characterization of the ecotin-pepsinogen fusion protein.

4.2.4 Large scale expression and purification of ecotin-pepsinogen

E. coli BL21(DE3)Gold harboring pEGP1 was grown in a fermentor in fed batch mode using a yeast extract enriched medium (Chapter 7.2). After induction with 1 mM IPTG at OD_{600} 18.7, the cells were grown for 5 h at 24 °C and were harvested at a final OD_{600} of 45 (605 g wet biomass from 9.5 L culture). The biomass was stored in aliquots at -80 °C.

As mentioned earlier, at the shake flask level, the expression of ecotin-pepsinogen was about 100 µg per liter OD1. Therefore, to purify ecotin-pepsinogen in mg quantity, large volumes of periplasmic extracts were required as a first step of purification. The periplasmic extract from 30 g biomass was prepared according to a protocol designed for frozen biomass (Rathore*, et al.*, 2003) but was additionally supplemented with 1 mM GSH and GSSG as redox shuffling agents (Malik*, et al.*, 2006). To reduce the viscosity of the extract, 6 µl benzonase and 6 mM $MgCl_2$ were added. To enrich His-tagged ecotin-pepsinogen from the periplasmic extract, a procedure was established using 40 ml Ni-NTA resin for 360 ml of periplasmic extract. Washing and elution was done in batch mode (Fig. 2, chapter 7.2). After Ni-NTA enrichment, 54.8% fusion protein (4.8 mg from 8.9 mg) was recovered. In the next step, anion exchange chromatography using HiTrapQ column was done at pH 7.5 with a yield of 46.8%. In the final purification using Superdex 75 column (Table 2), the final recovery of homogeneous purified ecotin-pepsinogen was 22.9% (Fig. 3A, chapter 7.2).

Table 2: Summary of the purification of ecotin-pepsinogen from the periplasmic extract

Step	Volume (ml)	Total protein (mg)[a]	Fusion protein (mg)[c]	Yield (%)
Periplasm preparation	360	612.6	8.9	100
Ni-NTA pool, dialyzed	190	47.0	4.8	54.8
HiTrapQ chromatography	9	11.5	4.1	46.8
Superdex 75 chromatography	22	5.2[b]	2.0	22.9

[a] Protein concentration determined by Bradford.
[b] Protein was determined at A_{280}, using a calculated ε = 83,240 mol^{-1} cm^{-1} for the ecotin-pepsinogen fusion protein.
[c] Protein concentration determined by pepsin assay, with 1 mg pepsin corresponding to 1.67 mg of the fusion protein.

Ecotin-pepsinogen was estimated using EGFP as a substrate (chapter 7.2) and porcine pepsinogen was served as a reference. Since the specific activity of enzymes from different sources and even isoforms of human pepsinogen showed variations towards different substrates and pH optima (Athauda*, et al.*, 1989), quantification of the homogeneous purified ecotin-pepsinogen was relied on the absorbance at 280 nm (Table 2).

When glutathione, a redox shuffling agent, was absent during the preparation of periplasmic extract, a 20 kDa protein coeluted along with ecotin fusion protein after gel filtration chromatography (Fig. 3B, chapter 7.2). N-terminal sequencing of this 20 kDa protein identified it as SlyD (a cytosolic PPIase) (Hottenrott*, et al.*, 1997). SlyD is a histidine-rich protein (Fig. 11) and purified via Ni-NTA chromatography (Hottenrott*, et al.*, 1997). The calculated pI of SlyD is close to ecotin-pepsinogen (4.86 and 4.71, respectively). That's why after anion exchange chromatography at pH 7.5, SlyD was coeluted. But after gel filtration chromatography, coelution of SlyD with ecotin-proinsulin indicated some kind of binding/association. One of the reasons may be that the periplasmic extraction was performed at pH 9.0 which is close to the pK_a of cysteines (~8.5). Another reason may be that the Ni^{2+} ions, facilitate metal catalyzed air oxidation during Ni-NTA chromatography (Dobeli*, et al.*, 1998). Therefore, disulfide bond shuffling between the ecotin fusion protein and SlyD which contains 6 cysteine residues distributed in the histidine rich patches might have occurred (Fig. 11). To prevent disulfide shuffling between the proteins, 1 mM reduced and oxidized glutathione was added in the periplasmic extraction buffer, keeping rest of the above mention purification step same. Under these conditions, the ecotin-pepsinogen obtained was free from SlyD contamination (Fig. 3A, chapter 7.2)

MK VAKDLVVS LAYQVRTEDG VLVDE S PVSA PLDYLHGHGS LI SG LE TALE
GH EVGD KFDV AVGA NDAYGQ YDENLV QRVP KDVFMGVDEL QVGMRFLAET
DQ GP V P VE IT AVEDDHVVVD GNHMLAGQNL KFNVEV VAIR EATEEELAHG
HVHGAHDHHH DHDHDGCCGG HGHDHGHEHG GEGCCGGKGN GGCGCH

Fig. 11. Primary sequence of SlyD (GenPept AAC41458)
The C-terminal part of SlyD is rich in histidine and cysteine residues and therefore binds tightly to divalent cations.

4.2.5 Characterization of ecotin-pepsinogen

The purified ecotin-pepsinogen was subjected to N-terminal sequencing which confirmed a proper cleavage of the ecotin signal sequence. Next, the purity and homogeneity of purified ecotin-pepsinogen was analyzed by RP-HPLC in the presence of 5 µM pepstatin, to prevent its autoactivation on a C4 column at low pH. When pepstatin was added, a single peak of ecotin-pepsinogen was obtained with a mass of 57,896 Da as determined by MALDI-TOF. The calculated reduced mass of ecotin-pepsinogen is 57,902 Da. After gel filtration chromatography using Superdex 200 column (Fig. 4, chapter 7.2), the apparent mass of ecotin-pepsinogen fusion protein was 136,000 Da, which corresponds to the calculated mass of the ecotin fusion protein in the dimeric state (115,804 Da).

The native folding of pepsinogen in the ecotin fusion state was evaluated by determination of the enzymatic activity of pepsin. Ecotin-pepsinogen was acidified at pH 2.0 for autoactivation of pepsinogen. Acidification of ecotin-pepsinogen leads to a fast activation (Fig. 5, chapter 7.2). Within 30 seconds of acidification, the full length ecotin-pepsinogen was converted into intermediate forms and a compact single band of pepsin appeared within 8 min of incubation. Human pepsinogen undergoes a sequential pathway upon acidification where the prosegment is removed in two steps (Athauda, *et al.*, 1989, Foltmann, 1988), via both uni- and bi-molecular mechanisms (Bustin and Conwayja.A, 1971, Marciniszyn, *et al.*, 1976). Ecotin is a very stable protein but also a good substrate for pepsin at low pH (Chung, *et al.*, 1983). Ecotin has 22 pepsin cleavage sites at pH >2.0. (http://www.expasy.ch/tools/peptidecutter/). As soon as pepsin appears in the acidified solution of ecotin-pepsinogen, it digests ecotin completely (Fig. 5, chapter 7.2) into very small fragments.

4.2.6 Future prospects of the ecotin fusion system

In future work, the production of ecotin-pepsinogen in the periplasmic space might be increased significantly if fermentation conditions, use of additives in the medium, periplasmic extraction conditions and purification are optimized. The properties of ecotin make it a promising periplasmic fusion tag. It is a relatively small size protein (16 kDa subunit), highly stable (withstands 100 °C and pH 1.0 for 30 min) and contains 1 disulfide bond in each subunit (Chung, *et al.*, 1983). Therefore, it undergoes a pathway of oxidative folding. Ecotin is made for the defence of *E. coli* against trypsin like serine proteases in the digestive tract and neutrphil elastase like serine proteases in the blood. No metabolic role or interaction of ecotin with other *E. coli* proteins was observed (Eggers, *et al.*, 2004). Ecotin is a dimeric protein and both C-termini protrude in opposite directions (Fig. 1A, chapter 7.2) which allow folding of passenger proteins at each end. Being a broad range serine protease inhibitor, affinity property of ecotin derivatives might be used for the purification. Protease binding sites of ecotin were already randomized (Stoop and Craik, 2003) to reduce the affinity to zymogens of serine proteases, which would help to elute proteins under mild conditions. Ecotin fusion protein can be quantified in a very sensitive trypsin inhibition assays (Kang, *et al.*, 2005). Ecotin is even stable and active in the cytosol; it can be used as cytoplasmic fusion tag (Kang, *et al.*, 2005 197). Ecotin can also be produced in monomeric native state after removal of the last 10 residues (Pal, *et al.*, 1996).

Pepsinogen contains 13 basic residues (Sogawa, *et al.*, 1983) and in general aspartic proteases contain 5-16 basic residues in the prosegment (Richter, *et al.*, 1998). Therefore, ecotin fusion to different aspartic proteases or other proteins rich in basic residues in the first 18 residues would improve translocation across cell membrane.

4.3 Periplasmic expression of human proinsulin with ecotin fusion

Is ecotin a suitable fusion tag for other classes of proteins? To address this important question, another protein, proinsulin, having three nonlinear disulfide bonds was chosen. It is different in size, folds and properties from pepsinogen. Since a self-cleavage of the fusion protein as with ecotin-pepsinogen does not occur for ecotin-proinsulin, an internal specific cleavage site and a histidine tag were inserted in the linker region of ecotin-proinsulin.

4.3.1 Cloning of ecotin-proinsulin

An ecotin-proinsulin fusion protein was designed to be expressed under control of the *trc* promoter in pTrc99a as shown in Fig. 1, chapter 7.3. Plasmid pTrc-EGP1 (Malik*, et al.*, 2006) was the source of ecotin gene and was transformed into *E. coli* GM2163 (*dam*¯ and *dcm*¯) strain. It was isolated and digested by *Bsp*EI, a methyl sensitive restriction enzyme, and *Sal*I. The digested vector contained the full-length ecotin plus three repeats of Gly-Ser $(GS)_3$ at the 3' end. The digested vector was dephosphorylated and purified by agarose gel electrophoresis. Proinsulin was amplified from pET20b-proinsulin (Schaffner*, et al.*, 2001). A long forward primer was designed to introduce three Gly-Ser repeats $(GS)_3$, a thrombin cleavage site (LVPRGS), a hexa-histidine tag $(H)_6$ and an arginine (R) residue just before the start of proinsulin with a unique cleavage site for *Bsp*EI. The reverse primer contained unique site for *Sal*I. The amplified proinsulin fragment was cloned into pCR-Blunt II-TOPO. After sequencing, this plasmid was transformed into *E. coli* GM2163 cells and was digested with *Bsp*EI and *Sal*I. The purified proinsulin insert was ligated into above mentioned vector. The resulting plasmid encoded ecotin followed by $(GS)_6$-LVPRGS-$(H)_6$-R and proinsulin (abbreviated as pEG-PI).

4.3.2 Expression, purification and characterization of ecotin-proinsulin

To purify and characterize the ecotin-proinsulin fusion protein, biomass of *E. coli* BL21(DE3)Gold expressing ecotin-proinsulin was prepared according to (Malik*, et al.*, 2006). Fermentation was done in fed batch mode. When glucose was depleted, feeding (30% yeast extract and 25% glycerol) was started and the temperature was lowered from 37 to 24 °C. After 1 hour of feeding, protein production was started at an OD_{600} of 31 by adding 1 mM IPTG. Cells

were harvested after 4 hours of incubation at OD_{600} of 59. A biomass of 909 g was obtained from 9.5 L culture and was stored in aliquots at -80 °C.

Periplasmic extract was prepared from the frozen biomass by osmotic shock according to (Malik, *et al.*, 2006). Activity of proinsulin was measured in a sandwich ELISA using monoclonal antibodies recognizing native folds of insulin and proinsulin (Winter, *et al.*, 2000). The yield of native ecotin-proinsulin in the periplasmic extract was about 2.0 µg per liter OD1. It is known that ecotin is acid stable (Chung, *et al.*, 1983) and insulin can be stabilized by acidification (Bryant, *et al.*, 1993). Therefore, the periplasmic extract was acidified and neutralized in order to precipitate acid labile proteins as a first purification step. To precipitate acid-labile proteins, 100 mM citric acid was added into the periplasmic extract for 10 min at room temperature. The precipitate was removed by centrifugation, followed by neutralization by adding 225 mM NaOH and 75 mM Tris-HCl, pH 8.5. The acidified and neutralized periplasmic extract was dialyzed against 50 mM Tris, 300 mM NaCl at pH 8.0 for Ni-NTA chromatography. The debris was removed by centrifugation at 90,000g for 30 min and the supernatant was filtered through a 0.2 µm filter. After the acid-treatment and neutralization of periplasmic extract, proinsulin corresponding to 0.23 µg of ecotin-proinsulin/ml extract was found. If an aliquot of the same extract without acid treatment was stored at 4 °C for 24 hours, no proinsulin was left, indicating proteolytic decay (data not shown).

Due to a large volume of periplasmic extract, a peristaltic pump was used for loading the solution onto an equilibrated Ni-NTA column. Washing and elution was done by FPLC with a gradient of 0-250 mM imidazole in 50 min at a flow rate of 1 ml/min. One peak with a shoulder appeared (Fig. 12), and the fractions (indicated with ↓) were analyzed by 15% SDS-PAGE (Fig. 13). A prominent band of 28 kDa was present in the shoulder of the peak while a band corresponding to nearly 20 kDa was present in the peak. N-terminal sequencing indicated that both proteins had the same N-terminal sequence (Ala-Glu-Ser-Val-Gln-Pro-) that started with the mature part of ecotin. It is well known that a periplasmic metalloprotease, pitrilysin, cleaves the oxidized B-chain of insulin at two points with an initial fast cleavage at Tyr16-Leu17 and a second slower cut at Phe25-Tyr26 (Cheng and Zipser, 1979). The calculated masses of intact and digested ecotin-proinsulin cleaved at Tyr16-Lue17, are 27,929.7 and 20,365.1 Da, respectively. It seems that the 20 kDa protein was a proteolytic fragment derived from full-length 28 kDa ecotin-

proinsulin. These results indicated that proinsulin was efficiently translocated to the periplasm (by ecotin), but undergoes severe proteolysis.

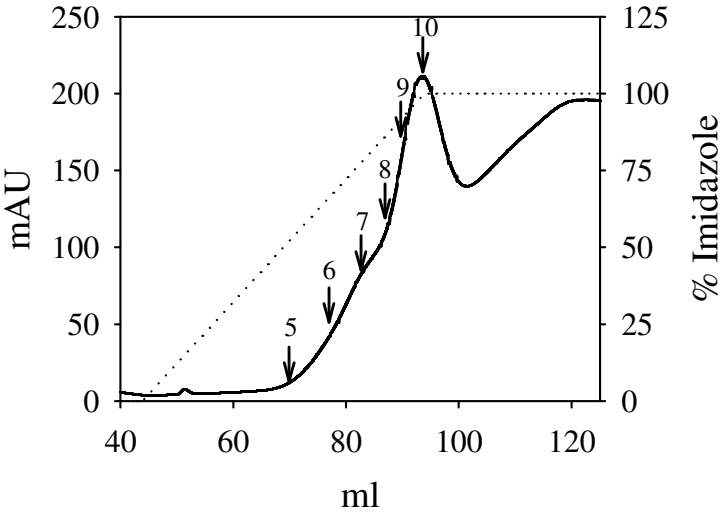


Fig. 12. Purification of ecotin-proinsulin from acidified and neutralized periplasmic extract via Ni-NTA chromatography.
The dotted line (….) shows a linear gradient of 0-250 mM imidazole and the solid line (—) represents the absorption of eluted protein at 280 nm. Arrows represent fractions that were used for ELISA and for SDS-PAGE analysis. Numbers above the arrows indicate the respective lane number on SDS-PAGE.

ELISA analysis of different fractions from the shoulder and from the peak showed that native proinsulin was present in the shoulder and in few fractions in the beginning of the peak (Fig. 12). The ELISA signal corresponded to the full length fusion protein visualized on SDS-PAGE (Fig. 13, lanes 5-8) while the digested proteins (Fig. 13, lanes 9-10) did not have any activity.

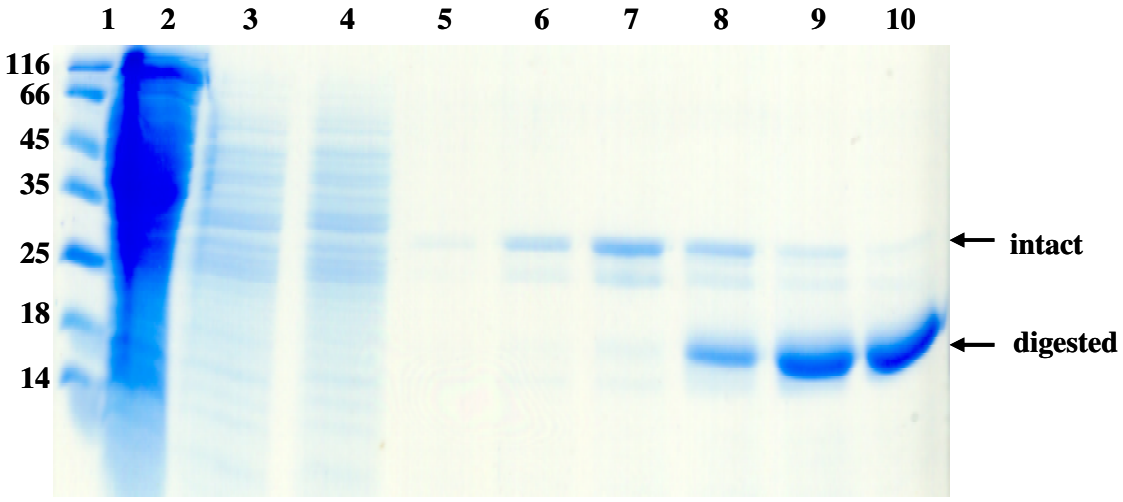


Fig. 13. Analysis on 16% SDS-PAGE after Ni-NTA chromatography.
Lane 1, marker proteins with the masses given in kDa; lane 2, periplasmic extract; lane 3, acidified-neutralized periplasmic extract; lane 4, flow-through from the Ni-NTA column, and lanes 5-10, eluted fractions after Ni-NTA, loaded sequentially as shown in above chromatogram. The arrows indicate the intact ecotin-proinsulin fusion protein (28 kDa) and digested (~20 kDa) fusion protein.

To further characterize the proinsulin, fractions from the Ni-NTA column showing proinsulin activity (Fig. 12) were pooled and digested with 0.75 mg thrombin at room temperature to separate proinsulin from the ecotin fusion. Ecotin is a broad range serine protease inhibitor but it inhibits thrombin activity by only 2% (Seymour, *et al.*, 1994). Ecotin is a homodimer (Chung, *et al.*, 1983) and the ecotin fusion protein is also dimeric (Malik, *et al.*, 2006). After thrombin digestion of ecotin-proinsulin, the signal obtained in the sandwich ELISA was doubled due to a monomerization of proinsulin from the dimeric ecotin-proinsulin fusion (data not shown).

Thrombin digested ecotin-proinsulin probes were subjected to RP-HPLC using a Nucleosil C18 column (Macherey-Nagel, Germany) (Winter, *et al.*, 2002). Two peaks were obtained after RP-HPLC (data not shown). Masses of 10,520.1 and 10,517.8 Da were determined for the major and minor peak, respectively by MALDI-TOF (REFLEX II; Bruker Daltonik GmbH, Bremen, Germany). The calculated value of the reduced His-tagged proinsulin is 10,517.8 Da (http://www.expasy.ch/cgi-bin/protparam). The difference might be explained by deamidation reaction which leads to increase in weight, 1 Da per deamidation. It is well known that under prolonged exposure to acidic pH, insulin undergoes deamidation (Brange, *et al.*, 1992) at positions Asn A21, Asn B3 and Gln B4 (Nilsson and Dobson, 2003). It seems that purified proinsulin was in the reduced and deamidated state, because the periplasm was extracted using a the protocol for frozen biomass at pH 9.0 (Rathore, *et al.*, 2003) in the presence of 1 mM redox shuffling agents (Malik, *et al.*, 2006). Therefore, to improve the yield and quality, and to avoid proteolysis of proinsulin, parameters were optimized for expression in high cell density fermentation, and the periplasm was extracted directly from fresh biomass at pH 8.0 (Winter, *et al.*, 2000) in the absence of redox-shuffling agents.

4.3.3 Optimization of ecotin-proinsulin expression

To increase the yield of ecotin-proinsulin, primary cultivation parameters such as inducer concentrations, antibiotic concentrations, post-induction incubation temperature and time were optimized. *E. coli* BL21(DE3)Gold harboring pEG-PI was grown in parallel fermentation vessels (Infors, Switzerland) in M9 medium (Jenzsch, *et al.*, 2006) at 37 °C containing 100 mg/L ampicillin. To optimize post-induction temperature and incubation time, after induction with 1 mM IPTG, the temperature was lowered to 35, 30, 25, and 20 °C, respectively and samples at

every hour were taken to extract the periplasmic content (Winter, *et al.*, 2000). As shown in Fig. 2A in chapter 7.3, 25-30 °C and 2-4 hours of incubation were the best post-induction temperature and time. To analyze the influence of the inducer on the yield of ecotin-proinsulin, 4 different concentrations of IPTG (0.05, 0.1, 0.5, and 1 mM) were added in parallel vessels and the temperature was lowered from 37 °C to 25 °C. The yield of ecotin-proinsulin was not much influenced by the inducer (Fig. 2B, chapter 7.3). To analyze the plasmid stability, 4 different concentration of ampicillin in combination with its acid stable analogue carbenicillin were tested. The effect of the antibiotic concentration on the yield of ecotin-proinsulin was negligible (Fig. 2C, chapter 7.3).

Next, high cell density fed batch fermentations were performed in a BBI Sartorius System's BIOSTAT® C 15-L- bioreactor with 8 L working volume. The batch process was started using the synthetic M9 or a semisynthetic (M9 + 1.5% peptone) medium. After depletion of glucose below 2 g/L, feeding (40% glucose solution) was started with a fixed exponential profile to obtain a specific biomass growth rate of 0.5/h. The protein production was induced with 1 mM IPTG once the cell density reached to a certain value, and the cultivation temperature was lowered to 25 °C. The exponential profile of the feeding rate was reduced to follow a specific biomass growth rate of 0.1/h.

Cultivation parameter and the yield obtained from the high cell density fermentations are summarized in table 1, chapter 7.3. The first fermentation was done using M9 medium and the protein synthesis was induced with 1 mM IPTG at a moderate optical density of 66. The culture was further grown for 3 hours at 25 °C. The volumetric yield of ecotin-proinsulin was 55 mg/L that corresponds to 18.5 mg of proinsulin per liter fermentation broth. In the next experiment, all the parameters were kept similar except that induction was done at a higher optical density of 158. Both, the specific and the volumetric yield of ecotin-proinsulin were reduced from 677 to 157 µg per liter OD1 and 55 to 27 mg/L, respectively.

The host strain used for expression contains the *relA* gene which produces ppGpp synthetase I during amino acid starvation (Jishage, *et al.*, 2002, Joseleau-Petit, *et al.*, 1999). RNA polymerase is a target for ppGpp, which plays a key role in the initiation of replication of several replicons. This stringent response is extremely important in plasmid replication (Barker, *et al.*, 2001). Most of the ColE1-like plasmids are unable to replicate efficiently in a *relA*$^+$ host during starvation for

different amino acids (Wrobel and Wegrzyn, 1998). Therefore, in the next fermentation experiment 1.5% peptone was added in M9 medium. After induction at high optical density of 161, the yield of fusion protein was strongly increased. The specific yield of ecotin-proinsulin increased nearly 6 fold from 157 to 908 µg per liter OD1 with a volumetric yield of 153 mg/L fermentation broth as compared to induction at similar optical density using only M9 medium. This experiment was repeated keeping all parameters constant. Now, the specific and volumetric yield of ecotin-proinsulin was 614 µg per liter OD1 and 111 mg/L, respectively.

4.3.4 Extraction of ecotin-proinsulin at large scale

To avoid cell lysis and reduce the release of cytosolic proteins including proteases in the periplasmic extract, fresh biomass immediately after completion of fed batch fermentation in semisynthetic M9 medium was used for the extraction of the periplasmic fraction. The osmotic shock procedure of Winter *et al* (Winter*, et al.*, 2000) was modified to scale up from milliliter to liter scale. The recovery of ecotin-proinsulin at liter scale was about 63% as compared to milliliter scale using the same biomass (70 mg and 111 mg per liter broth, respectively). The debris was removed by centrifugation and the supernatant was filtered through 0.2 µm filter. The periplasmic fraction was stored at -80 °C.

4.3.5 Affinity purification of ecotin-proinsulin

Due to a strong affinity of ecotin for trypsin as well as trypsinogen (Lengyel*, et al.*, 1998), 45 mg trypsinogen was coupled onto 5 ml NHS-activated Sepharose in a HiTrap HP column to capture ecotin-proinsulin from periplasmic extract. Periplasmic fraction, 250 ml, was loaded onto the column using a peristaltic pump at 4 °C. Bound protein was eluted by applying 50 mM HCl using a FPLC system and the fractions obtained were immediately neutralized by directly eluting in tubes containing 0.1 volume of 1 M Tris-HCl, pH 8.5. A sharp, symmetrical peak was obtained (Fig. 3A, chapter 7.3). The purity of eluted protein was analyzed on SDS-PAGE. A full-length ecotin-proinsulin (28 kDa) along with two smaller proteins (16 and 18 kDa) was eluted (Fig. 4A, lane 4, chapter 7.3). In the ELISA, eluted protein showed the presence of native proinsulin. After addition of thrombin in the pooled fractions of the peak, the proinsulin part was separated from the fusion protein, and the 28 kDa protein band was converted into the 10 kDa GS-$(H)_6$-R-proinsulin and the 18 kDa ecotin derivative (Fig. 4A, lane 5, chapter 7.3).

In the next step, the thrombin digested mixture was loaded onto a HisTrap FF column to separate GS-$(H)_6$-R-proinsulin from the remaining proteins (Fig. 3B, chapter 7.3). Wild-type endogenous ecotin and recombinant ecotin derivative were present in the flow through (Fig. 4A, lane 6, chapter 7.3). Two peaks eluted from the HisTrap column with a linear gradient of 0-300 mM imidazole (Fig. 3B, chapter 7.3). For the presence of proinsulin, both peaks were tested by ELISA. The proinsulin activity was nearly 10 fold less in the first peak as compared to the second peak (Table 2, chapter 7.3). For SDS-PAGE analysis, the protein from both peaks was 10 fold concentrated by TCA precipitation. In the first peak, no protein band was visible, while a single protein band was present in the second peak (Fig. 4B, lane 2-3, chapter 7.3). The molar extinction coefficient of ecotin-proinsulin dimer is quite high ($\varepsilon_{.}$ = 58,760 mol^{-1} cm^{-1}) as compared to proinsulin ($\varepsilon_{.}$ = 6,335 mol^{-1} cm^{-1}) (Pace*, et al.*, 1995). This might explain the relatively high absorption of thrombin undigested ecotin-proinsulin in the first peak after Ni-NTA and sharp elution due to presence of internal histidine-tag which might cause steric hindrance, while the absorption of second peak was low that contains GS-$(H)_6$-R-proinsulin and eluted broadly due to presence of terminal histidine-tag (Fig. 3B, chapter 7.3).

Affinity purification of ecotin-proinsulin is summarized in table 2, chapter 7.3. In the first step of purification, 250 ml of periplasmic extract from 4.5 g wet biomass containing 388 µg proinsulin was loaded on 45 mg immobilized trypsinogen column. Nearly 48% ecotin-proinsulin was captured. When the flow through of the trypsinogen column was passed second time to capture the rest of ecotin-proinsulin, no ecotin-proinsulin was eluted, indicating loss of ~52% proinsulin during the first passage through the trypsinogen column. In the next step, GS-$(H)_6$-R-proinsulin was purified from the thrombin digested mixture of ecotin-proinsulin on Ni-NTA chromatography with a yield of 43%. By these two steps, proinsulin was purified to homogeneity.

Actually, a commercial preparation of trypsinogen (Sigma) was immobilized directly without further purification. It might contain trypsin in small amounts. When thrombin-cleaved ecotin-proinsulin was passed through trypsinogen column to separate GS-$(H)_6$-R-proinsulin from the ecotin part, protein with a mass corresponding to insulin having insulin like folds (analyzed via ELISA) were present in the flow through, instead of GS-$(H)_6$-R-proinsulin. *In vitro*, trypsin removed major part of C-peptide from proinsulin and it became insulin like molecule (Kemmler*, et al.*, 1971). In future, better preparation of trypsinogen or inactive variants of trypsin could be

used for immobilization. The immobilized trypsinogen based purification strategy is an efficient and fast procedure, requiring no dialysis step. The ecotin-proinsulin purification strategy comprised of only two steps of affinity purification including one cleavage step and was completed within 5 hours, with 43% recovery of proinsulin in homogeneous form.

4.3.6 Characterization of ecotin-proinsulin

The protein probes obtained after trypsinogen column, thrombin digestion and Ni-NTA chromatography were characterized by ESI-MS (ESQUIRE-LC, Bruker Daltonik GmbH, Bremen, Germany). All probes were desalted by passing them through a $ZipTip_{C18}$ (Millipore). The probes were eluted with 4 µl 0.1% TFA in 50% acetonitrile. The mass spectroscopic data are shown in table 3, chapter 7.3. The full-length ecotin-proinsulin eluted from trypsinogen column had a mass of 27,921.0 Da which is 8.7 Da less than the calculated reduced mass of full length protein ecotin-proinsulin. Intact ecotin-proinsulin contains 4 disulfide bonds, the formation of each disulfide bond decreases 2 Da. Thus, the ecotin-proinsulin was obtained in its oxidized state. After thrombin digestion of ecotin-proinsulin, masses of 17,428 Da and 10,511 Da were obtained, corresponding to the oxidized ecotin part [ecotin-$(GS)_6$-LVPR, 17,428 Da] and oxidized proinsulin [GS-$(H)_6$R- proinsulin, 10,511 Da]. Wild-type ecotin which is naturally localized in the periplasm was also purified in the oxidized state (16,097 Da) along with ecotin-proinsulin via the trypsinogen column.

4.3.7 Production of ecotin-proinsulin in the protease deficient strain

As mentioned above, high level production of ecotin-proinsulin and its purification is a problem due to severe proteolysis. It seems that the 20 kDa proteolytic fragment (Fig. 13) was generated from ecotin-proinsulin due to proteolytic activity of pitrilysin at Tyr16-Lys17 or Phe25-Tyr26. Therefore, a comparative study of periplasmic expression was done in *E. coli* BL21(DE3)Gold (*lon*$^-$,*omp*T$^-$) and *E. coli* SF120 (*ptr*$^-$, *ompT*$^-$,*degP*$^-$) constructed by Meerman *et al.* (Meerman and Georgiou, 1994). The expression plasmid pEG-PI was transformed into *E. coli* BL21(DE3)Gold and *E. coli* SF120 with ampicillin as selection marker for *E. coli* BL21(DE3) and ampicillin, kanamycin and chloramphenicol for *E. coli* SF120. Both strains were inoculated into LB or M9 medium, with the respective antibiotics, and were grown overnight at 30 °C. Since the growth of SF120 strain was poor, higher volumes of overnight-grown cells were inoculated to attain similar

starting optical densities in fresh 20 ml medium containing the appropriate antibiotics. The cultures were grown at 30 °C to attain the mid-log phase. The protein synthesis was induced with 1 mM IPTG followed by lowering the temperature to 24 °C for 3 hours. Periplasmic extracts were prepared according to the method of Winter *et al.* (Winter*, et al.*, 2000), and native proinsulin was quantified by ELISA.

Growth of *E. coli* SF120 was apparently poor in both, LB and M9 medium, with a final optical density of 1.3 and 0.25, respectively. Under the same conditions, the optical density of *E. coli* BL21(DE3)Gold was 4, and 0.8, respectively. In *E. coli* SF120, 85 and 79 µg of fusion protein was produced per ml OD1 in LB and M9 medium, respectively. Thus, the specific yield was about 3-4 fold higher than that determined in *E. coli* BL21(DE3)Gold (21 and 23 µg, respectively; Fig. 5A, Chapter 7.3). However, due to the poor growth of SF120, the volumetric yield of ecotin-proinsulin was only marginally increased (Fig. 5B, chapter 7.3). Moreover, under high density fermentation conditions using SF120 strain, no ecotin-proinsulin fusion protein was detectable.

5 References

Alami, M., Luke, I., Deitermann, S., Eisner, G., Koch, H. G., Brunner, J. and Muller, M. (2003) Differential interactions between a twin-arginine signal peptide and its translocase in *Escherichia coli*. Mol. Cell. 12, 937-46

Allen, S. C., Barrett, C. M., Ray, N. and Robinson, C. (2002) Essential cytoplasmic domains in the *Escherichia coli* TatC protein. J. Biol. Chem. 277, 10362-6

Anson, M. L. (1938) The estimation of pepsin, trypsin, papain and cathepsin with hemoglobin. J.Gen. Physiol. 22, 65-78

Arie, J. P., Sassoon, N. and Betton, J. M. (2001) Chaperone function of FkpA, a heat shock prolyl isomerase, in the periplasm of *Escherichia coli*. Mol. Microbiol. 39, 199-210

Aristidou, A. A., San, K. Y. and Bennett, G. N. (1995) Metabolic engineering of *Escherichia coli* to enhance recombinant protein-production through acetate reduction. Biotechnol. Prog. 11, 475-78

Aslund, F., Ehn, B., Miranda-Vizuete, A., Pueyo, C. and Holmgren, A. (1994) Two additional glutaredoxins exist in *Escherichia coli*: glutaredoxin 3 is a hydrogen donor for ribonucleotide reductase in a thioredoxin/glutaredoxin 1 double mutant. Proc. Natl. Acad. Sci. USA. 91, 9813-7

Athauda, S. B. P., Tanji, M., Kageyama, T. and Takahashi, K. (1989) A comparative study on the NH_2-terminal amino acid sequences and some other properties of 6 Isozymic forms of human pepsinogens and pepsins. J. Biochem. 106, 920-27

Bader, M., Muse, W., Ballou, D. P., Gassner, C. and Bardwell, J. C. A. (1999) Oxidative protein folding is driven by the electron transport system. Cell. 98, 217-27

Bailey, J. E. (1993) Host-vector interactions in *Escherichia coli*. Adv. Biochem. Eng. Biotechnol. 48, 29-52

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

Baneyx, F. and Georgiou, G. (1992) Degradation of secreted proteins in *Escherichia coli*. Ann. N. Y. Acad Sci. 665, 301-8

Bardwell, J. C., Lee, J. O., Jander, G., Martin, N., Belin, D. and Beckwith, J. (1993) A pathway for disulfide bond formation in vivo. Proc. Natl. Acad. Sci. USA. 90, 1038-42

Bardwell, J. C., McGovern, K. and Beckwith, J. (1991) Identification of a protein required for disulfide bond formation in vivo. Cell. 67, 581-9

Barker, M. M., Gaal, T., Josaitis, C. A. and Gourse, R. L. (2001) Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro. J. Mol. Biol. 305, 673-88

Barrett, C. M., Ray, N., Thomas, J. D., Robinson, C. and Bolhuis, A. (2003) Quantitative export of a reporter protein, GFP, by the twin-arginine translocation pathway in *Escherichia coli*. Biochem. Biophys. Res. Commun. 304, 279-84

Bentley, W. E. and Kompala, D. S. (1990) Optimal induction of protein synthesis in recombinant bacterial cultures. Ann. N. Y. Acad. Sci. 589, 121-38

Berkmen, M., Boyd, D. and Beckwith, J. (2005) The nonconsecutive disulfide bond of *Escherichia coli* phytase (AppA) renders it dependent on the protein-disulfide isomerase, DsbC. J. Biol. Chem. 280, 11387-94

Berks, B. C., Sargent, F. and Palmer, T. (2000) The Tat protein export pathway. Mol. Microbiol. 35, 260-74

Bessette, P. H., Cotto, J. J., Gilbert, H. F. and Georgiou, G. (1999) In vivo and in vitro function of the *Escherichia coli* periplasmic cysteine oxidoreductase DsbG. J. Biol. Chem. 274, 7784-92

Betton, J. M. and Hofnung, M. (1996) Folding of a mutant maltose-binding protein of *Escherichia coli* which forms inclusion bodies. J. Biol. Chem. 271, 8046-52

Blackwell, J. R. and Horgan, R. (1991) A novel strategy for production of a highly expressed recombinant protein in an active form. FEBS Lett. 295, 10-2

Blondel, A., Nageotte, R. and Bedouelle, H. (1996) Destabilizing interactions between the partners of a bifunctional fusion protein. Prot. Eng. 9, 231-38

Bokman, S. H. and Ward, W. W. (1981) Renaturation of Aequorea green fluorescent protein. Biochem. Biophys. Res. Commun. 101, 1372-80

Bowers, C. W., Lau, F. and Silhavy, T. J. (2003) Secretion of LamB-LacZ by the signal recognition particle pathway of *Escherichia coli*. J. Bacteriol. 185, 5697-705

Brandts, J. F., Halvorson, H. R. and Brennan, M. (1975) Consideration of possibility that slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. Biochemistry. 14, 4953-63

Brange, J., Langkjaer, L., Havelund, S. and Volund, A. (1992) Chemical stability of insulin. 1. Hydrolytic degradation during storage of pharmaceutical preparations. Pharm. Res. 9, 715-26

Brosius, J., Erfle, M. and Storella, J. (1985) Spacing of the -10 and -35 Regions in the *tac* Promoter - Effect on its in vivo activity. J. Biol. Chem. 260, 3539-41

Bryant, C., Spencer, D. B., Miller, A., Bakaysa, D. L., Mccune, K. S., Maple, S. R., Pekar, A. H. and Brems, D. N. (1993) Acid Stabilization of Insulin. Biochemistry. 32, 8075-82

Bustin, M. and Conwayja.A (1971) Intramolecular Activation of Porcine Pepsinogen. J. Biol. Chem. 246, 615-&

Butt, T. R., Edavettal, S. C., Hall, J. P. and Mattern, M. R. (2005) SUMO fusion technology for difficult-to-express proteins. Prot. Expr. Purif. 43, 1-9

Carmel-Harel, O. and Storz, G. (2000) Roles of the glutathione- and thioredoxin-dependent reduction systems in the *Escherichia coli* and *Saccharomyces cerevisiae* responses to oxidative stress. Annu. Rev. Microbiol. 54, 439-61

Cedergren, L., Andersson, R., Jansson, B., Uhlen, M. and Nilsson, B. (1993) Mutational Analysis of the Interaction between *Staphylococcal* Protein-A and Human IgG. Prot. Eng. 6, 441-48

Chance, R. E., Kroeff, E. P., Hoffmann, J. A. and Frank, B. H. (1981) Chemical, physical, and biologic properties of biosynthetic human insulin. Diabetes Care. 4, 147-54

Chen, J., Song, J. L., Zhang, S., Wang, Y., Cui, D. F. and Wang, C. C. (1999) Chaperone activity of DsbC. J. Biol. Chem. 274, 19601-5

Cheng, Y. S. and Zipser, D. (1979) Purification and characterization of protease III from *Escherichia coli*. J. Biol. Chem. 254, 4698-706

Choi, J. H., Keum, K. C. and Lee, S. Y. (2006) Production of recombinant proteins by high cell density culture of Escherichia coli. Chem. Eng. Sci. 61, 876-85

Chou, M. M. and Kendall, D. A. (1990) Polymeric sequences reveal a functional interrelationship between hydrophobicity and length of signal peptides. J. Biol. Chem. 265, 2873-80

Chung, C. H., Ives, H. E., Almeda, S. and Goldberg, A. L. (1983) Purification from *Escherichia coli* of a periplasmic protein that is a potent inhibitor of pancreatic proteases. J. Biol. chem. 258, 11032-38

Clarke, J. and Fersht, A. R. (1993) Engineered disulfide bonds as probes of the folding pathway of Barnase - Increasing the stability of proteins against the rate of denaturation. Biochemistry. 32, 4322-29

Collinsracie, L. A., Mccolgan, J. M., Grant, K. L., Diblasiosmith, E. A., Mccoy, J. M. and Lavallie, E. R. (1995) Production of recombinant bovine enterokinase catalytic subunit in *Escherichia coli* using the novel secretory fusion partner DsbA. Bio/Technology. 13, 982-87

Cormack, B. P., Valdivia, R. H. and Falkow, S. (1996) FACS-optimized mutants of the green fluorescent protein (GFP). Gene. 173, 33-8

Cornishb.Aj and Knowles, J. R. (1969) pH-Dependence of pepsin-catalysed reactions. Biochem. J. 113, 353-&

Creagh, A. L., Ong, E., Jervis, E., Kilburn, D. G. and Haynes, C. A. (1996) Binding of the cellulose binding domain of exoglucanase Cex from *Cellulomonas fimi* to insoluble microcrystalline cellulose is entropically driven. Proc. Natl. Acad. Sci. USA. 93, 12229-34

Creighton, T. E. (1997a) How important is the molten globule for correct protein folding? Trends Biochem. Sci. 22, 6-10

Creighton, T. E. (1997b) Protein folding coupled to disulphide bond formation. Biol. Chem. 378, 731-44

da Silva Gomes, R. A., Batista, R. P., Costa de Almeida, A., da Fonseca, D. N., Juliano, L. and Hial, V. (2003) A fluorimetric method for the determination of pepsin activity. Anal. Biochem. 316, 11-4

Dalbey, R. E. and Vonheijne, G. (1992) Signal peptidases in prokaryotes and eukaryotes - a new protease family. Trends Biochem. Sci. 17, 474-78

Dartigalongue, C. and Raina, S. (1998) A new heat-shock gene, ppiD, encodes a peptidyl-prolyl isomerase required for folding of outer membrane proteins in *Escherichia coli*. EMBO J. 17, 3968-80

de Keyzer, J., van der Does, C., Kloosterman, T. G. and Driessen, A. J. (2003) Direct demonstration of ATP-dependent release of SecA from a translocating preprotein by surface plasmon resonance. J. Biol. Chem. 278, 29581-6

de Leeuw, E., Granjon, T., Porcelli, I., Alami, M., Carr, S. B., Muller, M., Sargent, F., Palmer, T. and Berks, B. C. (2002) Oligomeric properties and signal peptide binding by *Escherichia coli* Tat protein transport complexes. J. Mol. Biol. 322, 1135-46

de Leeuw, E., Poland, D., Mol, O., Sinning, I., ten Hagen-Jongman, C. M., Oudega, B. and Luirink, J. (1997) Membrane association of FtsY, the *E. coli* SRP receptor. FEBS Lett. 416, 225-9

DeLisa, M. P., Lee, P., Palmer, T. and Georgiou, G. (2004) Phage shock protein PspA of *Escherichia coli* relieves saturation of protein export via the Tat pathway. J. Bacteriol. 186, 366-73

Dieci, G., Bottarelli, L., Ballabeni, A. and Ottonello, S. (2000) tRNA-assisted overproduction of eukaryotic ribosomal proteins. Prot. Expr. Purif. 18, 346-54

Dobeli, H., Andres, H., Breyer, N., Draeger, N., Sizmann, D., Zuber, M. T., Weinert, B. and Wipf, B. (1998) Recombinant fusion proteins for the industrial production of disulfide bridge containing peptides: Purification, oxidation without concatamer formation, and selective cleavage. Prot. Expr. Purif. 12, 404-14

Driessen, A. J., Fekkes, P. and van der Wolk, J. P. (1998) The Sec system. Curr. Opin. Microbiol. 1, 216-22

Dubochet, J., Mcdowall, A. W., Menge, B., Schmid, E. N. and Lickfeld, K. G. (1983) Electron-microscopy of frozen-hydrated bacteria. J. Bacteriol. 155, 381-90

Duguay, A. R. and Silhavy, T. J. (2004) Quality control in the bacterial periplasm. Biochem. Biophys. Acta. 1694, 121-34

Duplay, P., Bedouelle, H., Fowler, A., Zabin, I., Saurin, W. and Hofnung, M. (1984) Sequences of the MalE Gene and of its product, the Maltose-Binding Protein of *Escherichia coli* K12. J. Biol. Chem. 259, 606-13

Eggers, C. T., Murray, L. A., Delmar, V. A., Day, A. G. and Craik, C. S. (2004) The periplasmic serine protease inhibitor ecotin protects bacteria against neutrophil elastase. Biochem. J. 379, 107-18

Englander, S. W. and Kallenbach, N. R. (1983) Hydrogen-exchange and structural dynamics of proteins and nucleic-acids. Q. Rev. Biophys. 16, 521-655

Fekkes, P. and Driessen, A. J. (1999) Protein targeting to the bacterial cytoplasmic membrane. Microbiol. Mol. Biol. Rev. 63, 161-73

Flower, A. M., Doebele, R. C. and Silhavy, T. J. (1994) PrlA and PrlG suppressors reduce the requirement for signal sequence recognition. J. Bacteriol. 176, 5607-14

Foltmann, B. (1988) Activation of Human Pepsinogens. FEBS Lett. 241, 69-72

Freedman, R. B. (1989) Protein disulfide isomerase: multiple roles in the modification of nascent secretory proteins. Cell. 57, 1069-72

Fuh, G., Mulkerrin, M. G., Bass, S., McFarland, N., Brochier, M., Bourell, J. H., Light, D. R. and Wells, J. A. (1990) The human growth hormone receptor. Secretion from *Escherichia coli* and disulfide bonding pattern of the extracellular binding domain. J. Biol. Chem. 265, 3111-5

Gennity, J., Goldstein, J. and Inouye, M. (1990) Signal peptide mutants of *Escherichia coli*. J. Bioenerg. Biomembr. 22, 233-69

Gilkes, N. R., Jervis, E., Henrissat, B., Tekant, B., Miller, R. C., Jr., Warren, R. A. and Kilburn, D. G. (1992) The adsorption of a bacterial cellulase and its two isolated domains to crystalline cellulose. J. Biol. Chem. 267, 6743-9

Gilkes, N. R., Warren, R. A., Miller, R. C., Jr. and Kilburn, D. G. (1988) Precise excision of the cellulose binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. J. Biol. Chem. 263, 10401-7

Gothel, S. F. and Marahiel, M. A. (1999) Peptidyl-prolyl cis-trans isomerases, a superfamily of ubiquitous folding catalysts. Cell Mol. Life Sci. 55, 423-36

Gottesman, S. (1996) Proteases and their targets in *Escherichia coli*. Annu. Rev. Genet. 30, 465-506

Greenwood, J. M., Gilkes, N. R., Kilburn, D. G., Miller, R. C. and Warren, R. A. J. (1989) Fusion to an endoglucanase allows alkaline phosphatase to bind to cellulose. FEBS Lett. 244, 127-31

Greenwood, J. M., Ong, E., Gilkes, N. R., Warren, R. A. J., Miller, R. C. and Kilburn, D. G. (1992) Cellulose binding domains - Potential for purification of complex proteins. Prot. Eng. 5, 361-65

Gruninger-Leitch, F., Schlatter, D., Kung, E., Nelbock, P. and Dobeli, H. (2002) Substrate and inhibitor profile of BACE (beta-secretase) and comparison with other mammalian aspartic proteases. J. Biol. Chem. 277, 4687-93

Gu, S. Q., Peske, F., Wieden, H. J., Rodnina, M. V. and Wintermeyer, W. (2003) The signal recognition particle binds to protein L23 at the peptide exit of the *Escherichia coli* ribosome. RNA. 9, 566-73

Guddat, L. W., Bardwell, J. C. and Martin, J. L. (1998) Crystal structures of reduced and oxidized DsbA: investigation of domain motion and thiolate stabilization. Structure. 6, 757-67

Guzman, L. M., Belin, D., Carson, M. J. and Beckwith, J. (1995) Tight Regulation, modulation, and high-level expression by vectors containing the arabinose pBAD promoter. J. Bacteriol. 177, 4121-30

Hammarberg, B., Nygren, P. A., Holmgren, E., Elmblad, A., Tally, M., Hellman, U., Moks, T. and Uhlen, M. (1989) Dual affinity fusion approach and its use to express recombinant human insulin-like growth factor-II. Proc. Natl. Acad. Sci. USA. 86, 4367-71

Heras, B., Edeling, M. A., Schirra, H. J., Raina, S. and Martin, J. L. (2004) Crystal structures of the DsbG disulfide isomerase reveal an unstable disulfide. Proc. Natl. Acad. Sci. USA. 101, 8876-81

Herskovits, A. A., Bochkareva, E. S. and Bibi, E. (2000) New prospects in studying the bacterial signal recognition particle pathway. Mol. Microbiol. 38, 927-39

Hiniker, A. and Bardwell, J. C. (2004) In vivo substrate specificity of periplasmic disulfide oxidoreductases. J. Biol. Chem. 279, 12967-73

Hobot, J. A., Carlemalm, E., Villiger, W. and Kellenberger, E. (1984) Periplasmic gel - new concept resulting from the reinvestigation of bacterial-cell envelope Ultrastructure by new methods. J. Bacteriol. 160, 143-52

Horne, S. M. and Young, K. D. (1995) *Escherichia coli* and other species of the Enterobacteriaceae encode a protein similar to the family of Mip-like FK506 binding proteins. Arch. Microbiol. 163, 357-65

Hottenrott, S., Schumann, T., Pluckthun, A., Fischer, G. and Rahfeld, J. U. (1997) The *Escherichia coli* SlyD is a metal ion-regulated peptidyl-prolyl cis/trans-isomerase. J. Biol. Chem. 272, 15697-701

Hwang, C., Sinskey, A. J. and Lodish, H. F. (1992) Oxidized redox state of glutathione in the endoplasmic reticulum. Science. 257, 1496-502

Inouye, K. and Fruton, J. S. (1967) Studies on the specificity of pepsin. Biochemistry. 6, 1765-77

Izard, J. W. and Kendall, D. A. (1994) Signal peptides: exquisitely designed transport promoters. Mol. Microbiol. 13, 765-73

Jander, G., Martin, N. L. and Beckwith, J. (1994) Two cysteines in each periplasmic domain of the membrane protein DsbB are required for its function in protein disulfide bond formation. EMBO J. 13, 5121-7

Jensen, E. B. and Carlsen, S. (1990) Production of recombinant human growth hormone in *Escherichia coli* - Expression of different precursors and physiological effects of glucose, acetate, and salts. Biotechnol. Bioeng. 36, 1-11

Jenzsch, M., Gnoth, S., Kleinschmidt, M., Simutis, R. and Lubbert, A. (2006) Improving the batch-to-batch reproducibility in microbial cultures during recombinant protein production by guiding the process along a predefined total biomass profile. Bioprocess Biosyst. Eng. 29, 315-21

Jishage, M., Kvint, K., Shingler, V. and Nystrom, T. (2002) Regulation of sigma factor competition by the alarmone ppGpp. Genes Dev. 16, 1260-70

Joly, J. C., Leung, W. S. and Swartz, J. R. (1998) Overexpression of *Escherichia coli* oxidoreductases increases recombinant insulin-like growth factor-I accumulation. Proc. Natl. Acad. Sci. USA. 95, 2773-77

Jonasson, P., Liljeqvist, S., Nygren, P. A. and Stahl, S. (2002) Genetic design for facilitated production and recovery of recombinant proteins in *Escherichia coli*. Biotechnol. Appl. Biochem. 35, 91-105

Jones, L. J., Upson, R. H., Haugland, R. P., Panchuk-Voloshina, N. and Zhou, M. (1997) Quenched BODIPY dye-labeled casein substrates for the assay of protease activity by direct fluorescence measurement. Anal. Biochem. 251, 144-52

Joseleau-Petit, D., Vinella, D. and D'Ari, R. (1999) Metabolic alarms and cell division in *Escherichia coli*. J. Bacteriol. 181, 9-14

Kadokura, H., Katzen, F. and Beckwith, J. (2003) Protein disulfide bond formation in prokaryotes. Annu. Rev. Biochem. 72, 111-35

Kajava, A. V., Zolov, S. N., Kalinin, A. E. and Nesmeyanova, M. A. (2000) The net charge of the first 18 residues of the mature sequence affects protein translocation across the cytoplasmic membrane of gram-negative bacteria. J. Bacteriol. 182, 2163-9

Kamitani, S., Akiyama, Y. and Ito, K. (1992) Identification and characterization of an *Escherichia coli* gene required for the formation of correctly folded alkaline phosphatase, a periplasmic enzyme. EMBO J. 11, 57-62

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

Kane, J. F., Violand, B. N., Curran, D. F., Staten, N. R., Duffin, K. L. and Bogosian, G. (1992) Novel in-frame two codon translational hop during synthesis of bovine placental lactogen in a recombinant strain of *Escherichia coli*. Nucleic Acids Res. 20, 6707-12

Kang, S. H., Park, J. J., Chung, S. S., Bang, O. S. and Chung, C. H. (2005) Strategies for assaying deubiquitinating enzymes. Methods Enzymol. 398, 500-08

Kebir, M. O. and Kendall, D. A. (2002) SecA specificity for different signal peptides. Biochemistry. 41, 5573-80

Kemmler, W., Peterson, J. D. and Steiner, D. F. (1971) Studies on the conversion of proinsulin to insulin. I. Conversion in vitro with trypsin and carboxypeptidase B. J. Biol. Chem. 246, 6786-91

Khokhlova, O. V. and Nesmeianova, M. A. (2003) Interaction of SecB and SecA with the N-terminal region of mature alkaline phosphatase on its secretion in *Escherichia coli*. Mol Biol (Mosk). 37, 712-8

Lang, K., Schmid, F. X. and Fischer, G. (1987) Catalysis of protein folding by prolyl isomerase. Nature. 329, 268-70

Larsson, M., Brundell, E., Nordfors, L., Hoog, C., Uhlen, M. and Stahl, S. (1996) A general bacterial expression system for functional analysis of cDNA-encoded proteins. Prot. Expr. Purif. 7, 447-57

Leduc, M., Frehel, C. and Vanheijenoort, J. (1985) Correlation between degradation and ultrastructure of peptidoglycan during autolysis of *Escherichia coli*. J. Bacteriol. 161, 627-35

Lee, C., Li, P., Inouye, H., Brickman, E. R. and Beckwith, J. (1989) Genetic studies on the inability of Beta-Galactosidase to be translocated across the *Escherichia coli* cytoplasmic membrane. J. Bacteriol. 171, 4609-16

Lee, H. C. and Bernstein, H. D. (2001) The targeting pathway of *Escherichia coli* presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal. Proc. Natl. Acad. Sci. USA. 98, 3471-6

Lee, S. Y. (1996) High cell-density culture of *Escherichia coli*. Trends Biotechnol. 14, 98-105

Lengyel, Z., Pal, G. and Sahin-Toth, M. (1998) Affinity purification of recombinant trypsinogen using immobilized ecotin. Prot. Expr. Purif. 12, 291-4

Li, X., Zhang, G., Ngo, N., Zhao, X., Kain, S. R. and Huang, C. C. (1997) Deletions of the *Aequorea victoria* green fluorescent protein define the minimal domain required for fluorescence. J. Biol. Chem. 272, 28545-9

Liang, S. M., Allet, B., Rose, K., Hirschi, M., Liang, C. M. and Thatcher, D. R. (1985) Characterization of human interleukin 2 derived from *Escherichia coli*. Biochem. J. 229, 429-39

Lin, X. L., Wong, R. N. and Tang, J. (1989) Synthesis, purification, and active site mutagenesis of recombinant porcine pepsinogen. J. Biol. Chem. 264, 4482-9

Lipinska, B., Zylicz, M. and Georgopoulos, C. (1990) The HtrA (DegP) protein, essential for *Escherichia coli* survival at high temperatures, is an endopeptidase. J. Bacteriol. 172, 1791-7

Liu, J. and Walsh, C. T. (1990) Peptidyl-prolyl cis-trans-isomerase from *Escherichia coli*: a periplasmic homolog of cyclophilin that is not inhibited by cyclosporin A. Proc. Natl. Acad. Sci. USA. 87, 4028-32

Luirink, J. and Dobberstein, B. (1994) Mammalian and *Escherichia coli* signal recognition particles. Mol. Microbiol. 11, 9-13

Luirink, J., von Heijne, G., Houben, E. and de Gier, J. W. (2005) Biogenesis of inner membrane proteins in *Escherichia coli*. Annu. Rev. Microbiol. 59, 329-55

Luli, G. W. and Strohl, W. R. (1990) Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. Appl. Environ. Microbiol. 56, 1004-11

Maier, R., Eckert, B., Scholz, C., Lilie, H. and Schmid, F. X. (2003) Interaction of trigger factor with the ribosome. J. Mol. Biol. 326, 585-92

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

Malik, A., Rudolph, R. and Sohling, B. (2006) A novel fusion protein system for the production of native human pepsinogen in the bacterial periplasm. Prot. Expr. Purif. 47, 662-71

Mallick, P., Boutz, D. R., Eisenberg, D. and Yeates, T. O. (2002) Genomic evidence that the intracellular proteins of archaeal microbes contain disulfide bonds. Proc. Natl. Acad. Sci. USA. 99, 9679-84

Marciniszyn, J., Huang, J. S., Hartsuck, J. A. and Tang, J. (1976) Mechanism of intramolecular activation of pepsinogen - Evidence for an intermediate-delta and involvement of active site of pepsin in intramolecular activation of pepsinogen. J. Biol. Chem. 251, 7095-102

Martin, J. L., Bardwell, J. C. and Kuriyan, J. (1993) Crystal structure of the DsbA protein required for disulphide bond formation in vivo. Nature. 365, 464-8

Martsev, S. P., Tsybovsky, Y. I., Stremovskiy, O. A., Odintsov, S. G., Balandin, T. G., Arosio, P., Kravchuk, Z. I. and Deyev, S. M. (2004) Fusion of the antiferritin antibody V_L domain to barnase results in enhanced solubility and altered pH stability. Prot. Eng. Des. Select. 17, 85-93

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

McCarthy, A. A., Haebel, P. W., Torronen, A., Rybin, V., Baker, E. N. and Metcalf, P. (2000) Crystal structure of the protein disulfide bond isomerase, DsbC, from *Escherichia coli*. Nat. Struct. Biol. 7, 196-9

McNulty, D. E., Claffee, B. A., Huddleston, M. J., Porter, M. L., Cavnar, K. M. and Kane, J. F. (2003) Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli*. Prot. Expr. Purif. 27, 365-74

McPhie, P. (1976) A turbidimetric milk-clotting assay for pepsin. Anal. Biochem. 73, 258-61

Meerman, H. J. and Georgiou, G. (1994) Construction and characterization of a set of *E. coli* strains deficient in all known loci affecting the proteolytic stability of secreted recombinant proteins. Bio/Technology . 12, 1107-10

Mergulhao, F. J., Summers, D. K. and Monteiro, G. A. (2005) Recombinant protein secretion in *Escherichia coli*. Biotechnol. Adv. 23, 177-202

Messens, J. and Collet, J. F. (2006) Pathways of disulfide bond formation in *Escherichia coli*. Int. J. Biochem. Cell Biol. 38, 1050-62

Missiakas, D., Betton, J. M. and Raina, S. (1996) New components of protein folding in extracytoplasmic compartments of *Escherichia coli* SurA, FkpA and Skp/OmpH. Mol. Microbiol. 21, 871-84

Mogensen, J. E. and Otzen, D. E. (2005) Interactions between folding factors and bacterial outer membrane proteins. Mol. Microbiol. 57, 326-46

Muller, M., Koch, H. G., Beck, K. and Schaefer, U. (2001) Protein traffic in bacteria: Multiple routes from the ribosome to and across the membrane. Prog. Nucl. Acid Res. Mol. Biol. 66, 107-57

Nagai, K., Oubridge, C., Kuglstatter, A., Menichelli, E., Isel, C. and Jovine, L. (2003) Structure, function and evolution of the signal recognition particle. EMBO J. 22, 3479-85

Nallamsetty, S., Austin, B. P., Penrose, K. J. and Waugh, D. S. (2005) Gateway vectors for the production of combinatorially-tagged His(6)-MBP fusion proteins in the cytoplasm and periplasm of *Escherichia coli*. Prot. Sci. 14, 2964-71

Nelson, J. W. and Creighton, T. E. (1994) Reactivity and ionization of the active site cysteine residues of DsbA, a protein required for disulfide bond formation in vivo. Biochemistry. 33, 5974-83

Nilsson, B., Moks, T., Jansson, B., Abrahmsen, L., Elmblad, A., Holmgren, E., Henrichson, C., Jones, T. A. and Uhlen, M. (1987) A Synthetic IgG-binding domain based on Staphylococcal Protein-A. Prot. Eng. 1, 107-13

Nilsson, J., Jonasson, P., Samuelsson, E., Stahl, S. and Uhlen, M. (1996) Integrated production of human insulin and its C-peptide. J. Biotechnol. 48, 241-50

Nilsson, J., Nilsson, P., Williams, Y., Pettersson, L., Uhlen, M. and Nygren, P. A. (1994) Competitive elution of Protein-A fusion proteins allows specific recovery under mild conditions. Eur. J. Biochem. 224, 103-08

Nilsson, J., Stahl, S., Lundeberg, J., Uhlen, M. and Nygren, P. A. (1997) Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins. Prot. Expr. Purif. 11, 1-16

Nilsson, M. R. and Dobson, C. M. (2003) Chemical modification of insulin in amyloid fibrils. Prot. Sci. 12, 2637-41

Nishiyama, K., Fukuda, A., Morita, K. and Tokuda, H. (1999) Membrane deinsertion of SecA underlying proton motive force-dependent stimulation of protein translocation. EMBO J. 18, 1049-58

Nomine, Y., Ristriani, T., Laurent, C., Lefevre, J. F., Weiss, E. and Trave, G. (2001) Formation of soluble inclusion bodies by HPV E6 oncoprotein fused to maltose-binding protein. Prot. Expr. Purif. 23, 22-32

Northrop, J. H. (1930) Crystalline pepsin: I. Isolation and test of purity. J.Gen. Physiol. 13, 739-66

Novokhatny, V. and Ingham, K. (1997) Thermodynamics of maltose binding protein unfolding. Prot. Sci. 6, 141-46

Nygren, P. A., Eliasson, M., Abrahmsen, L., Uhlen, M. and Palmcrantz, E. (1988) Analysis and use of the serum albumin binding domains of streptococcal protein-G. J. Mol. Recognit. 1, 69-74

Oliver, D., Norman, J. and Sarker, S. (1998) Regulation of *Escherichia coli* secA by cellular protein secretion proficiency requires an intact gene X signal sequence and an active translocon. J. Bacteriol. 180, 5240-42

Ong, E., Gilkes, N. R., Miller, R. C., Warren, R. A. J. and Kilburn, D. G. (1991) Enzyme immobilization using a cellulose-binding domain - Properties of a beta-Glucosidase fusion protein. Enzy. Microbial. Technol. 13, 59-65

Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. Prot. Sci. 4, 2411-23

Pal, G., Szilagyi, L. and Graf, L. (1996) Stable monomeric form of an originally dimeric serine proteinase inhibitor, ecotin, was constructed via site directed mutagenesis. FEBS Lett. 385, 165-70

Palmer, T. and Berks, B. C. (2003) Moving folded proteins across the bacterial cell membrane. Microbiology. 149, 547-56

Pantoliano, M. W., Ladner, R. C., Bryan, P. N., Rollence, M. L., Wood, J. F. and Poulos, T. L. (1987) Protein engineering of Subtilisin BPN' - Enhanced stabilization through the introduction of 2 cysteines to form a disulfide bond. Biochemistry. 26, 2077-82

Patterson, G., Day, R. N. and Piston, D. (2001) Fluorescent protein spectra. J. Cell Sci. 114, 837-38

Pines, O. and Inouye, M. (1999) Expression and secretion of proteins in *E. coli*. Mol. Biotechnol. 12, 25-34

Planson, A. G., Guijarro, J. I., Goldberg, M. E. and Chaffotte, A. F. (2003) Assistance of maltose binding protein to the in vivo folding of the disulfide-rich C-terminal fragment from *Plasmodium falciparum* merozoite surface protein 1 expressed in *Escherichia coli*. Biochemistry. 42, 13202-11

Prinz, W. A., Spiess, C., Ehrmann, M., Schierle, C. and Beckwith, J. (1996) Targeting of signal sequenceless proteins for export in *Escherichia coli* with altered protein translocase. EMBO J. 15, 5209-17

Qi, H. Y. and Bernstein, H. D. (1999) SecA is required for the insertion of inner membrane proteins targeted by the *Escherichia coli* signal recognition particle. J. Biol. Chem. 274, 8993-97

Rao, B. D. N., Kemple, M. D. and Prendergast, F. G. (1980) Proton nuclear magnetic resonance and fluorescence spectroscopic studies of segmental mobility in aequorin and a green fluorescent protein from *Aequorea forskalea*. Biophys. J. 32, 630-32

Rathore, A. S., Bilbrey, R. E. and Steinmeyer, D. E. (2003) Optimization of an osmotic shock procedure for isolation of a protein product expressed in *E. coli*. Biotechnol. Prog. 19, 1541-6

Richarme, G. and Caldas, T. D. (1997) Chaperone properties of the bacterial periplasmic substrate binding proteins. J. Biol. Chem. 272, 15607-12

Richter, C., Tanaka, T. and Yada, R. Y. (1998) Mechanism of activation of the gastric aspartic proteinases: pepsinogen, progastricsin and prochymosin. Biochem. J. 335, 481-90

Rietsch, A., Bessette, P., Georgiou, G. and Beckwith, J. (1997) Reduction of the periplasmic disulfide bond isomerase, DsbC, occurs by passage of electrons from cytoplasmic thioredoxin. J. Bacteriol. 179, 6602-8

Rouviere, P. E. and Gross, C. A. (1996) SurA, a periplasmic protein with peptidyl-prolyl isomerase activity, participates in the assembly of outer membrane porins. Genes Dev. 10, 3170-82

Russel, M. (1995) Thioredoxin genetics. Methods Enzymol. 252, 264-74

Sachdev, D. and Chirgwin, J. M. (1998) Order of fusions between bacterial and mammalian proteins can determine solubility in *Escherichia coli*. Biochem. Biophys. Res. Commun. 244, 933-37

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

San, K. Y., Bennett, G. N., Aristidou, A. A. and Chou, C. H. (1994) Strategies in high level expression of recombinant protein in *Escherichia coli*. Recombinant DNA Technology Ii. 721, 257-67

Santini, C. L., Ize, B., Chanal, A., Muller, M., Giordano, G. and Wu, L. F. (1998) A novel sec-independent periplasmic protein translocation pathway in *Escherichia coli*. EMBO J. 17, 101-12

Sarker, S., Rudd, K. E. and Oliver, D. (2000) Revised translation start site for secM defines an atypical signal peptide that regulates *Escherichia coli* secA expression. J. Bacteriol. 182, 5592-95

Schafer, U., Beck, K. and Muller, M. (1999) Skp, a molecular chaperone of gram-negative bacteria, is required for the formation of soluble periplasmic intermediates of outer membrane proteins. J. Biol. Chem. 274, 24567-74

Schaffner, J., Winter, J., Rudolph, R. and Schwarz, E. (2001) Cosecretion of chaperones and low-molecular-size medium additives increases the yield of recombinant disulfide-bridged proteins. Appl. Environ. Microbiol. 67, 3994-4000

Schierle, C. F., Berkmen, M., Huber, D., Kumamoto, C., Boyd, D. and Beckwith, J. (2003) The DsbA signal sequence directs efficient, cotranslational export of passenger proteins to the *Escherichia coli* periplasm via the signal recognition particle pathway. J. Bacteriol. 185, 5706-13

Schirra, H. J., Renner, C., Czisch, M., Huber-Wunderlich, M., Holak, T. A. and Glockshuber, R. (1998) Structure of reduced DsbA from *Escherichia coli* in solution. Biochemistry. 37, 6263-76

Schmoldt, H. U., Wentzel, A., Becker, S. and Kolmar, H. (2005) A fusion protein system for the recombinant production of short disulfide bond rich cystine knot peptides using barnase as a purification handle. Prot. Expr. Purif. 39, 82-89

Schnell, D. J. and Hebert, D. N. (2003) Protein translocons: multifunctional mediators of protein translocation across membranes. Cell. 112, 491-505

Scotti, P. A., Valent, Q. A., Manting, E. H., Urbanus, M. L., Driessen, A. J., Oudega, B. and Luirink, J. (1999) SecA is not required for signal recognition particle-mediated targeting and initial membrane insertion of a nascent inner membrane protein. J. Biol. Chem. 274, 29883-8

Sevier, C. S. and Kaiser, C. A. (2002) Formation and transfer of disulphide bonds in living cells. Nature Rev. Mol. Cell Biol. 3, 836-47

Seymour, J. L., Lindquist, R. N., Dennis, M. S., Moffat, B., Yansura, D., Reilly, D., Wessinger, M. E. and Lazarus, R. A. (1994) Ecotin is a potent anticoagulant and reversible tight-binding inhibitor of factor Xa. Biochemistry. 33, 3949-58

Shao, F., Bader, M. W., Jakob, U. and Bardwell, J. C. (2000) DsbG, a protein disulfide isomerase with chaperone activity. J. Biol. Chem. 275, 13349-52

Shaw, A. and Bott, R. (1996) Engineering enzymes for stability. Curr. Opin. Struct. Biol. 6, 546-50

Sogawa, K., Fujiikuriyama, Y., Mizukami, Y., Ichihara, Y. and Takahashi, K. (1983) Primary structure of human pepsinogen gene. J. Biol. Chem. 258, 5306-11

Spiess, C., Beil, A. and Ehrmann, M. (1999) A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. Cell. 97, 339-47

Stahl, S., Sjolander, A., Nygren, P. A., Berzins, K., Perlmann, P. and Uhlen, M. (1989) A dual expression system for the generation, analysis and purification of antibodies to a repeated sequence of the *Plasmodium falciparum* Antigen Pf155/Resa. J. Immunol. Methods. 124, 43-52

Stanley, N. R., Palmer, T. and Berks, B. C. (2000) The twin arginine consensus motif of Tat signal peptides is involved in Sec-independent protein targeting in *Escherichia coli*. J. Biol. Chem. 275, 11591-96

Stewart, E. J., Katzen, F. and Beckwith, J. (1999) Six conserved cysteines of the membrane protein DsbD are required for the transfer of electrons from the cytoplasm to the periplasm of *Escherichia coli*. EMBO J. 18, 5963-71

Stock, J. B., Rauch, B. and Roseman, S. (1977) Periplasmic space in *Salmonella typhimurium* and *Escherichia coli*. J. Biol. Chem. 252, 7850-61

Stoop, A. A. and Craik, C. S. (2003) Engineering of a macromolecular scaffold to develop specific protease inhibitors. Nature Biotechnol. 21, 1063-68

Strauch, K. L., Johnson, K. and Beckwith, J. (1989) Characterization of degP, a gene required for proteolysis in the cell envelope and essential for growth of *Escherichia coli* at high temperature. J. Bacteriol. 171, 2689-96

Studier, F. W. and Moffatt, B. A. (1986) Use of Bacteriophage T7 RNA-polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. 189, 113-30

Takahashi, Y. H., Inaba, K. and Ito, K. (2004) Characterization of the menaquinone-dependent disulfide bond formation pathway of *Escherichia coli*. J. Biol. Chem. 279, 47057-65

Thiry, M. and Cingolani, D. (2002) Optimizing scale-up fermentation processes. Trends Biotechnol. 20, 103-05

Trepod, C. M. and Mott, J. E. (2002) A spontaneous runaway vector for production scale expression of bovine somatotropin from *Escherichia coli*. Appl. Microbiol. Biotechnol. 58, 84-8

van Straaten, M., Missiakas, D., Raina, S. and Darby, N. J. (1998) The functional properties of DsbG, a thiol-disulfide oxidoreductase from the periplasm of *Escherichia coli*. FEBS Lett. 428, 255-8

Vanwielink, J. E. and Duine, J. A. (1990) How Big Is the Periplasmic Space. Trends Biochem. Sci. 15, 136-37

Voet, D. and Voet, J. G. (1995) Biochemistry, Second edition, John Wiley & Sons, Chapter 1, p14

von Heijne, G. (1985) Signal sequences. The limits of variation. J. Mol. Biol. 184, 99-105

Vonheijne, G. (1984) How signal sequences maintain cleavage specificity. J. Mol. Biol. 173, 243-51

Ward, W. W. and Bokman, S. H. (1982) Reversible denaturation of *Aequorea* green fluorescent protein: physical separation and characterization of the renatured protein. Biochemistry. 21, 4535-40

Warren, R. A., Beck, C. F., Gilkes, N. R., Kilburn, D. G., Langsford, M. L., Miller, R. C., Jr., O'Neill, G. P., Scheufens, M. and Wong, W. K. (1986) Sequence conservation and region shuffling in an endoglucanase and an exoglucanase from *Cellulomonas fimi*. Proteins. 1, 335-41

Winter, J., Lilie, H. and Rudolph, R. (2002) Renaturation of human proinsulin - a study on refolding and conversion to insulin. Anal. Biochem. 310, 148-55

Winter, J., Neubauer, P., Glockshuber, R. and Rudolph, R. (2000) Increased production of human proinsulin in the periplasmic space of *Escherichia coli* by fusion to DsbA. J. Biotechnol. 84, 175-85

Woodbury, R. L., Topping, T. B., Diamond, D. L., Suciu, D., Kumamoto, C. A., Hardy, S. J. S. and Randall, L. L. (2000) Complexes between protein export chaperone SecB and SecA - Evidence for separate sites on SecA providing binding energy and regulatory interactions. J. Biol. Chem. 275, 24191-98

Wrobel, B. and Wegrzyn, G. (1998) Replication regulation of ColE1-like plasmids in amino acid-starved *Escherichia coli*. Plasmid. 39, 48-62

Wunderlich, M., Otto, A., Seckler, R. and Glockshuber, R. (1993) Bacterial protein disulfide isomerase - Efficient catalysis of oxidative protein folding at acidic pH. Biochemistry. 32, 12251-56

Zapun, A., Bardwell, J. C. and Creighton, T. E. (1993) The reactive and destabilizing disulfide bond of DsbA, a protein required for protein disulfide bond formation in vivo. Biochemistry. 32, 5083-92

Zapun, A. and Creighton, T. E. (1994) Effects of DsbA on the disulfide folding of bovine pancreatic trypsin inhibitor and alpha-lactalbumin. Biochemistry. 33, 5202-11

Zheng, M., Aslund, F. and Storz, G. (1998) Activation of the OxyR transcription factor by reversible disulfide bond formation. Science. 279, 1718-21

6 Abbreviations

A_{450}	absorbance at 450 nanometer
Ala	alanine
Arg	arginine
Asn	asparagine
ATP	adenosine 5'-triphosphate
BPTI	bovine pancreatic trypsin inhibitor
cm^{-1}	per centimeter
Cys	cysteine
°C	degree Celsius
Da	Dalton
DNA	deoxyribonucleic acid
Dsb	disulfide bonding protein
E_0'	standard redox potential
E. coli	*Escherichia coli*
EGFP	enhanced green fluorescent protein
ELISA	enzyme-linked immunosorbent assay
ESI-MS	electron spray ionization-mass spectroscopy
FPLC	fast protein liquid chromatography
g	gram
GFP	green fluorescent protein
Gln	glutamine
Glu	glutamic acid
Gly	glycine
GS	glycine-serine
GSH	reduced glutathione
GSSG	oxidized glutathione
h	hour
h^{-1}	per hour
H^+	proton
His	histidine
Hsp	heat shock protein
Ig	immunoglobulin
IGF-I	insulin-like growth factor-I
IPTG	isopropyl-b-D-thiogalactopyranoside
kcal	kilocalory
kDa	kilodalton
K_m	Michaelis constant
L	liter
LB	Luria Bertani
Lue	leucine
LVPR	leucine-valine-proline-arginine
Lys	lysine
MALDI-TOF	matrix assisted laser desorption/ionization- time-of-flight
mg	milligram
min	minute

ml	milliliter
mM	millimolar
mV	millivolt
µg	microgram
µl	microliter
µM	micromolar
µm	micrometer
mol^{-1}	per mole
mRNA	messenger ribonucleic acid
NADPH	nicotinamide adenine dinucleotide phosphate, reduced form
ng	nanogram
NHS	N-hydroxysuccinimide
Ni	nickel
nm	nanometer
nM	nanomolar
NTA	nitriloacetic acid
OD_{600}	optical density at 600 nanometer
ompT	signal sequence of outer membrane protease T
PAGE	polyacrylamide gel electrophoresis
pelB	signal sequence of pectate lyase
pg	picogram
pI	isoelectric point
ppGpp	guanosine-tetraphosphate
PMF	proton motive force
Pro	proline
PVDF	polyvinylidene fluoride
RNA	ribonucleic acid
RP-HPLC	reversed-phase high performance liquid chromatography
SDS	Sodium dodecyl sulfate
Ser	serine
SRP	signal recognition particle
S-S	disulfide bond
TAT	twin arginine translocation
TCA	trichloroacetic acid
TFA	trifluoroacetic acid
T_m	melting temperature
TNBSA	2,4,6-Trinitrobenzene Sulfonic Acid
tRNA	transfer RNA
Trp	tryptophan
Trx	Thioredoxin
Tyr	tyrosine
Val	valine
V_L	light chain variable

7 Published and submitted manuscripts and patents

7.1 Malik, A., Rudolph, R., Söhling, B. (2005) Use of enhanced green fluorescent protein to determine pepsin at high sensitivity. *Analytical Biochemistry*, 340, 252–258.

7.2 Malik, A., Rudolph, R., Söhling, B. (2006) A novel fusion protein system for the production of native human pepsinogen in the bacterial periplasm Protein Expr. Purif., 47, 662–671

7.3 Malik, A., Jenzsch, M., Lübbert, A., Rudolph, R., Söhling, B.(2006) Periplasmic production of native human proinsulin as a fusion to *E. coli* ecotin (under revision-Protein Expr. Purif.)

7.4 Malik, A. Söhling, B. & Rudolph, R. (2005) - "Periplasmatische Produktion therapeutisch relevanter Proteine durch Fusion an *Escherichia coli* Ecotin "- Deutsche Patent- und Markenamt (Patent pending-10 2005 050 732.8-41)

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Analytical Biochemistry 340 (2005) 252–258
ANALYTICAL BIOCHEMISTRY
www.elsevier.com/locate/yabio

Use of enhanced green fluorescent protein to determine pepsin at high sensitivity

Ajamaluddin Malik, Rainer Rudolph, Brigitte Söhling *

Institut für Biotechnologie, Martin-Luther Universität Halle, Kurt-Mothes-Str. 3, 06120 Halle, Germany

Received 4 November 2004
Available online 10 March 2005

Abstract

A fluorometric assay for pepsin and pepsinogen was developed using enhanced green fluorescent protein (EGFP) as a substrate. Acid denaturation of EGFP resulted in a complete loss of fluorescence that was completely reversible on neutralization. In the proteolytic assay procedure, acid-denatured EGFP was digested by pepsin or activated pepsinogen. After neutralization, the remaining amount of undigested EGFP refolded and was determined by fluorescence. Under standard digestion conditions, 4.8–24.0 ng pepsin or pepsinogen was used. Using porcine pepsin as a standard, 38 ± 6.7 ng EGFP was digested per min^{-1} ng pepsin^{-1}. Activated porcine pepsinogen revealed a similar digestion rate (37.2 ± 5.2 ng EGFP min^{-1} ng activated pepsinogen^{-1}). The sensitivity of the proteolysis assay depended on the time of digestion and the temperature. Increasing temperature and incubation time allowed quantification of pepsin or pepsinogen in a sample even in the picogram range. The pepsin-catalyzed EGFP digestion showed typical Michaelis–Menten kinetics. K_m and V_{max} values were determined for the pepsin and activated pepsinogen. Digestion of EGFP by pepsin revealed distinct cleavage sites, as analyzed by SDS–PAGE.

Keywords: Enhanced green fluorescent protein; Pepsin; Pepsinogen; Aspartic protease

Pepsin is the principle acidic protease present in the stomach. It is secreted as a zymogen (pepsinogen) in the gastric chief cells of fundic mucosa, where it is autocatalytically converted into its active form under acidic conditions. According to its physiological role, it has a broad substrate specificity but preferentially cleaves its substrates at large hydrophobic residues [1,2]. Pepsin is a member of the large family of aspartic proteases. These enzymes are of great medical and pharmaceutical interest because some of them play important roles in the development of various human diseases (e.g., hypertension), in the formation of gastric ulcers, in HIV viral maturation, and as a prognostic tool for breast tumor invasiveness [3].

Since the time when pepsin was first crystallized [4], a number of different substrates and assay systems have been established to detect pepsin and its proform in biological samples. Determination of proteolytic activity was usually done using cattle hemoglobin as substrate with a subsequent quantification of the peptides released [5–7], now most convenient by A_{280}. Gelatin, casein, edestin, and milk proteins have also been tested as substrates [7], but hemoglobin still remains the preferred substrate. It is suitable for samples containing 1–10 μg pepsin ml^{-1}. The milk-clotting assay is more sensitive, but it is not precise and selective because turbidity changes of a milk solution can be evoked by a variety of proteases and factors [8].

Enzyme–substrate interactions and substrate specificity of pepsin have been analyzed using synthetic peptides containing a defined cleavage position (two aromatic amino acids). These peptides contain chromogenic

* Corresponding author. Fax: +49 345 5527013.
E-mail address: brigitte.soehling@biochemtech.uni-halle.de (B. Söhling).

0003-2697/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ab.2005.02.022

compounds such as a nitrophenyl residue or *N*-acetyl-diiodotyrosine [9–14]. More recently, an intramolecularly quenched fluorogenic peptide was used as a substrate [15]. The peptide-based assays are highly specific and sensitive, but a drawback of peptide substrates is the low stability and solubility [12], often leading to only small changes in the overall absorbance. Thus, pepsin activity in complex samples can be masked by a high protein background.

Fluorescence analysis is very sensitive, and pepsin determination using a fluorescent substrate should be very promising, in particular, if the substrate is highly soluble and stable. Green fluorescent protein (GFP, 27 kDa)[1] is an acidic and compact globular protein as well as a popular research tool in cellular and developmental biology [16,17]. Enhanced green fluorescent protein (EGFP) is a variant that carries two missense mutations (F64L and S65T) that increase fluorescence intensity 35-fold over that of wild-type GFP [18]. Deletion mapping experiments have shown that nearly the entire protein sequence is required for chromophore formation (7–229 of 238) [19]. EGFP fluorescence is more sensitive to acidic denaturation than is GFP [20]. As will be shown here, EGFP fluorescence can be completely restored by neutralization, and it is directly proportional to the amount of native protein. The aim of this study was to explore the use of EGFP as a substrate for pepsin and activated pepsinogen to determine these proteins at high sensitivity. An assay system was established and is described below.

Materials and methods

Materials

Porcine pepsin, porcine pepsinogen, and pepstatin A were purchased from Sigma. Benzonase (grade II) was obtained from Merck, and lysozyme was obtained from Roche. Ni–NTA resin was purchased from Novagen, and Hi-Trap Q HP prepacked column was purchased from Amersham. All other chemicals were of reagent grade. Molecular mass standard was purchased from Peqlab (Erlangen, Germany). Fluorescence measurements were taken with a Hitachi-F4500 fluorescence spectrophotometer.

EGFP expression and purification

pET19b-EGFP was constructed using the EGFP sequence from pEGFP-N1 (Clontech) and pET19b (Novagen) as expression vector (Esser and Böhm, unpublished). The resulting plasmid encoded EGFP N-terminally fused to a His_{10} tag. For protein production, *Escherichia coli* BL21(DE3) Gold (Stratagene) harboring pET19b-EGFP was used. The cells were grown at 37 °C in 1 L Luria-Bertani (LB) broth containing ampicillin (100 mg L^{-1}). After the addition of 1 mM isoproylthiogalactoside (IPTG, $OD_{600} \sim 1.5$) and incubation for 5 h, the cells were harvested by centrifugation at 6000 rpm for 10 min and were resuspended in 30 ml of resuspension buffer (0.05 M Tris–HCl, 0.3 M NaCl, 0.02 M imidazole, 2 mM $MgCl_2$, pH 8.0). Lysozyme (6 mg) was added, and the suspension was incubated at room temperature for 30 min. Then 1.25 µl benzonase (500 U/ml) was added, and the cells were disrupted in a Gaulin French press (7 passages, 550 bar). After incubation at room temperature for 30 min, the solution was centrifuged at 30,000 rpm at 4 °C for 1 h to discard cell debris. The supernatant was passed through a 0.2-µM membrane and was applied to an 8-ml Ni–NTA column (fast protein liquid chromatography, FPLC) previously equilibrated with 32 ml of 0.05 M Tris–HCl, 0.3 M NaCl, and 0.02 M imidazole (pH 8.0). After washing with 45 ml of the same buffer, the His-tagged EGFP was eluted with 0.05 M Tris–HCl, 0.3 M NaCl, and 0.25 M imidazole (pH 8.0) at a flow rate of 1 ml min^{-1}. EGFP-containing fractions were pooled and dialyzed against 0.05 M Tris–HCl and 0.01 M NaCl (pH 8.0).

For ion exchange chromatography, a 5-ml Hi-Trap Q HP column was used. The column was equilibrated with 25 ml of 0.05 M Tris–HCl and 0.01 M NaCl (pH 8.0) (buffer A) and loaded with the Ni–NTA-purified EGFP solution. After washing with 65 ml of buffer A, the elution was performed by 80 ml of 0–20% buffer B (0.05 M Tris–HCl, 2 mM EDTA, 2 M NaCl, pH 8.0), and 10 ml of 20–100% buffer B, at a flow rate of 2 ml min^{-1}. Homogeneity of the purified protein was analyzed by SDS–PAGE and by reversed-phase HPLC on a C4 column (Vydac). During reversed-phase HPLC, a single peak (29,556 Da) was obtained. The mass was determined by matrix-assisted laser desorption/ionization time of flight (MALDI–TOF) on a REFLEX II (Bruker Daltonik, Bremen, Germany). The yield of purified EGFP was approximately 7.5 mg g^{-1} biomass. The protein solution (1.1 mg ml^{-1}) was stored at −20 °C in 0.05 M Tris–HCl, 2 mM EDTA, and 0.06 M NaCl (pH 8.0).

Denaturation and renaturation of EGFP

A fresh working solution of EGFP was prepared (0.11 mg ml^{-1} in 10 mM Tris–HCl, pH 7.5) and was diluted to give 0–55 µg EGFP in a total volume of 500 µl. Acidification was done by the addition of 0.1 volume of 1 M citric acid (pH 2.0) and incubation at 20 °C for 10 min. Renaturation of EGFP was achieved by adding 400 µl of 1 M Tris–HCl (pH 8.5). The final pH of the

[1] *Abbreviations used:* GFP, green fluorescent protein; EGFP, enhanced green fluorescent protein; LB, Luria—Bertani; IPTG, isoproylthiogalactoside; FPLC, fast protein liquid chromatography; MALDI–TOF, matrix-assisted laser desorption/ionization time of flight.

solution was 8.0. Fluorescence of renatured EGFP was determined using a Hitachi F-4500 fluorescence spectrophotometer with excitation at 490 nm. Emission was monitored from 500 to 530 nm, with a maximum at 508 nm. Untreated EGFP in the same concentrations served as a control.

EGFP proteolysis assay

The standard proteolysis assay was performed as follows. EGFP (11 μg) was denatured according to the procedure described above. Then 50 μl protease solution (pepsin or activated pepsinogen) dissolved in 0.1 M citric acid (pH 2.0) was added to 550 μl denatured EGFP, and the reaction was incubated at 20 °C for 10 min. After the addition of 400 μl of 1 M Tris–HCl (pH 8.5) and incubation at 20 °C for 30 min, the fluorescence of the solution was monitored as above. A standard curve using acid-treated but undigested EGFP served as a control. For higher sensitivity, digestion was done at 37 °C for 1 or 3 h. Inhibition of proteolysis was done with 5 μM pepstatin A added to the acid-denatured EGFP prior to the addition of pepsin and pepsinogen. Determination of time-dependent proteolysis was performed under standard conditions in the presence of 55 μg EGFP. Kinetic measurements were done under the standard conditions, with substrate concentrations varying from 0.88 to 22 μg (29.7–734.2 nM). For these data, digestion was stopped after 15 min.

Protein masses and extinction coefficients

EGFP concentrations were calculated using an extinction coefficient of 55,000 [21] and a mass of 29,600 Da. For porcine pepsin (34,600 Da) and porcine pepsinogen (39,600 Da), extinction coefficients of 51,715 and 53,205, respectively, were calculated according to [22].

Activation of pepsinogen

Pepsinogen activation was done by incubation of porcine pepsinogen (125.8 $\mu g\,ml^{-1}$) in 0.01 M Tris–HCl (pH 7.5) with 0.1 volume of 1 M citric buffer (pH 2.0) at 37 °C. Complete cleavage of the zymogen would produce active pepsin at a final concentration of 100 $\mu g\,ml^{-1}$. The amount of active enzyme at various time intervals was determined by the hemoglobin assay method at pH 2.0. Incubation for 10 min at 37 °C proved to be sufficient for 100% activation.

Hemoglobin proteolysis assay

A solution of hemoglobin (2%w/v in 60 mM HCl) was prepared. The solution was equilibrated at 37 °C for 10 min, and 0.5 ml was combined with 0.1 ml protease solution (pepsin or activated pepsinogen). After 10 min incubation at 37 °C, 1 ml of 5%(w/v) trichloroacetic acid was added and the tubes were shaken for 5 min at 37 °C. Then the probes were centrifuged at 13,000 rpm for 15 min at room temperature, and the amount of soluble peptides in the supernatant was monitored at 280 nm. For activation analysis of pepsinogen, an equivalent amount of pepsin was used as a control.

SDS–PAGE analysis of EGFP proteolysis

EGFP was prepared in 0.1 M citric acid (pH 2.0) at 55 $\mu g\,ml^{-1}$. The protein solution was incubated with 23.8 $ng\,ml^{-1}$ pepsin, or 24.0 $ng\,ml^{-1}$ pepsinogen, at 20 °C. After 0, 2, 6, 10, 20, and 30 min, 600 μl was removed and the cleavage was stopped by adding 400 μl of 1 M Tris–HCl (pH 8.5). Each sample (1 ml) was precipitated by the addition of 250 μl of 100% trichloroacetic acid. The sample was finally dissolved in 200 μl of 0.1 M Tris–HCl (pH 8.0). Each time, 1.1 μg precipitated protein (4 μl) was loaded onto a 17.5% SDS polyacrylamide gel. The gel was stained with silver nitrate.

Results and discussion

EGFP was expressed and purified from *E. coli* as described in Materials and methods. A standard solution of the purified protein was diluted to monitor its fluorescence at different protein concentrations (Fig. 1). The EGFP fluorescence increase was linearly dependent on the amount of protein present, and a calibration curve for native fluorescent EGFP up to a concentration of 55 $\mu g\,ml^{-1}$ was obtained.

Use of EGFP as a substrate for pepsin would require acid denaturation, followed by a neutralization step to refold the remaining undigested substrate molecules. To analyze EGFP fluorescence after acidification and subsequent renaturation, varying amounts of EGFP (1.1–55.0 μg) were acidified to pH 2.0 for 10 min and then neutralized again to pH 8.0. EGFP fluorescence obtained after this treatment was the same as that of the untreated sample, indicating that EGFP could completely refold under these conditions (Fig. 1). Recovery of EGFP fluorescence proceeded very rapidly, as shown in Fig. 1B. Renaturation was completed within 5 min at room temperature, and the fluorescence was stable for at least 1 h.

Recovery of wild-type GFP following acid, base, or guanidine denaturation has been reported only up to 90% [23], in contrast to the nearly complete regain of fluorescence after acid treatment of EGFP shown here. GFP variants with improved in vivo maturation properties at 37 °C (so-called "folding mutants") have been described and classified according to the locations of the respective mutations, with one class having mutations

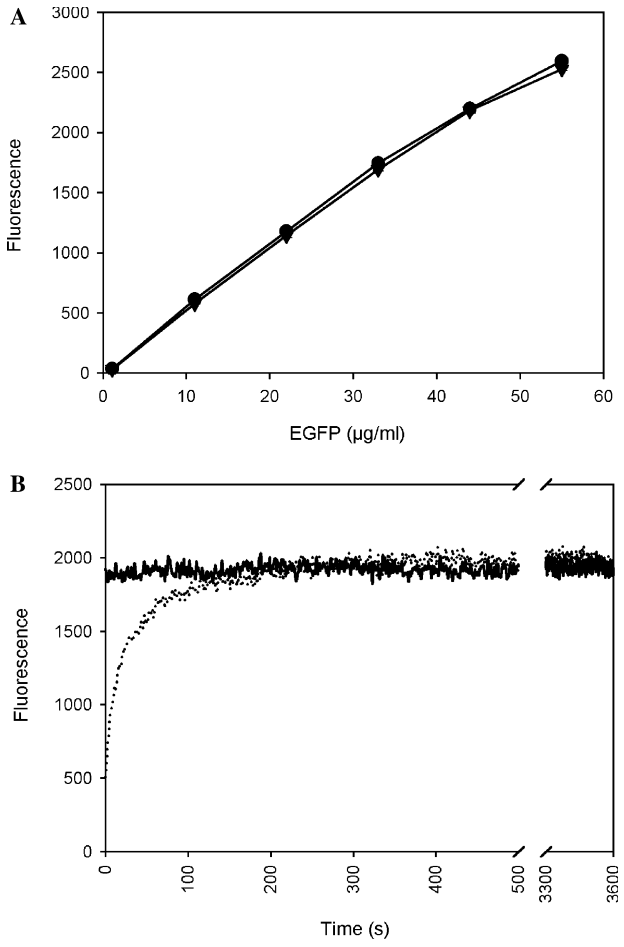


Fig. 1. Fluorescence analysis of native and renatured EGFP. (A) Fluorescence analysis of EGFP at different protein concentrations. Native EGFP (●) was determined in 0.4 M Tris–HCl (pH 8.5), renatured EGFP (▼) was determined after acid denaturation (addition of 0.1 volume of 1 M citric acid, pH 2.0, 10 min), and subsequent neutralization (addition of 0.4 volume of 1 M Tris–HCl, pH 8.5, 30 min). The final pH was 8.0. Excitation was done at 490 nm (slit 2.5 nm), and emission was monitored at 508 nm (slit 5.0 nm). (B) Time-dependent renaturation of acid-denatured EGFP. Fluorescence of acid-denatured and neutralized EGFP (11 µg ml^{-1}) is shown as a dotted line. Fluorescence of native EGFP (11 µg ml^{-1}, solid line) served as a control.

within or close to the chromophore [24], fitting well with our data. Quite recently, detailed GFP refolding studies after acid denaturation were published [25].

Next, to monitor pepsin-catalyzed digestion of EGFP, the acidified substrate was treated with different amounts of protease. After neutralization, undigested EGFP molecules were determined by fluorescence. EGFP fluorescence decreased in a linear way that directly corresponded to the amount of protease added (Fig. 2). In the presence of pepstatin A (5 µM), the cleavage was completely inhibited (data not shown). Depending on the digestion temperature and length, the sensitivity of the assay could be varied from a nanogram scale to a picogram scale. For standard assay conditions, a digestion time of 30 min at 20 °C was chosen to detect 5–25 ng pepsin. After 3 h digestion at 37 °C, pepsin con-

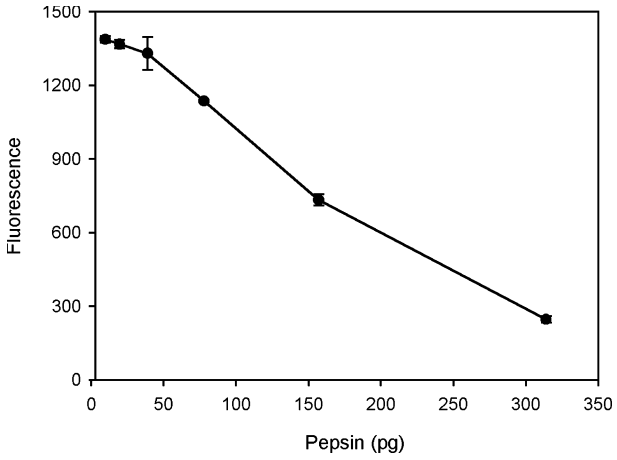


Fig. 2. EGFP cleavage using porcine pepsin. EGFP (11 µg in 10 mM Tris, pH 7.5) was acid-denatured by the addition of 0.1 volume of 1 M citric acid (pH 2.0) and was digested by serially diluted samples of porcine pepsin in 0.1 M citric acid (pH 2.0). After 3 h incubation at 37 °C, the reaction was stopped by the addition of 0.4 volume of 1 M Tris–HCl (pH 8.5), and EGFP fluorescence was determined.

centrations as low as 50 pg were detectable (Fig. 2). After incubation for 1 h at 37 °C, 3 ng pepsin was sufficient to digest up to 11 µg EGFP.

Time-dependent cleavage of EGFP in the presence of different amounts of pepsin is shown in Fig. 3. The digestion rate was directly dependent on the amount of protease added, as is typical for an enzyme-catalyzed reaction. Thus, in the presence of 4.8, 14.3, and 23.8 ng pepsin, similar specific digestion rates were obtained (38 ± 6.7 ng EGFP ng pepsin^{-1} min^{-1}).

Pepsinogen is the zymogen form of pepsin. It is inactive at neutral pH but undergoes self-cleavage to pepsin in the acidic environment of the gastric juice. Pepsinogen activation proceeds by an autocatalytic intramolecular mechanism within a few seconds, depending on acidity, temperature, and pepsinogen concentration [26,27]. Prior to the use of pepsinogen in our assay system, we routinely activated the protein by adding 0.1 volume of 1 M citric acid (pH 2.0) and incubation for 10 min at 37 °C (see Materials and methods). Pepsinogen was completely converted into pepsin under these conditions.

EGFP proteolysis using activated pepsinogen samples was analyzed as before. As with pepsin, time-dependent EGFP proteolysis depended on the amount of activated pepsinogen (Fig. 3B). The specific digestion rates obtained with 4.8, 14.4, and 24.0 ng pepsinogen were quite similar to those obtained with the mature enzyme (37.2 ± 5.2 ng EGFP min^{-1} ng pepsin^{-1}).

Thus, the EGFP-based assay is suited to determine pepsinogen and pepsin in a nanogram range, and this is more sensitive than most of the other systems described so far. Digestion of the intramolecularly quenched fluorogenic peptide substrate was done in the presence of 8–10 µg pepsin, but this assay might be more sensitive [15].

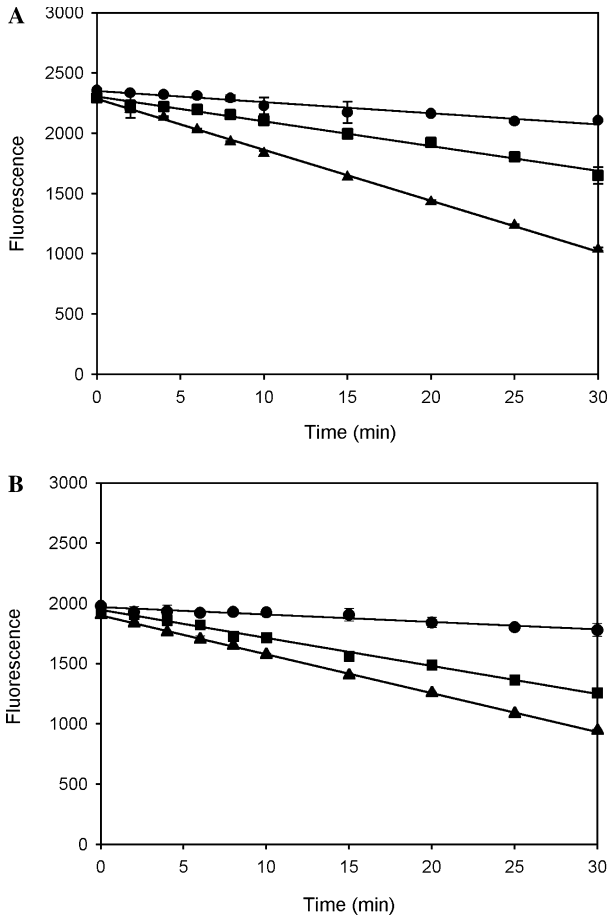


Fig. 3. Time-dependent cleavage of EGFP. (A) Acid-denatured EGFP (55 μg) was incubated with 4.8 ng (●), 14.3 ng (■), and 23.8 ng (▲) pepsin at 20 °C. At the indicated times, the reaction was stopped by the addition of 0.4 volume of 1 M Tris–HCl (pH 8.5) and EGFP fluorescence was determined. A specific digestion rate of 38.2 ± 6.7 ng EGFP min^{-1} ng pepsin^{-1} was obtained. (B) Acid-denatured EGFP (55 μg) was incubated with 4.8 (●), 14.4 ng (■), and 24.0 ng (▲) of activated pepsinogen at 20 °C. At the indicated times, the reaction was stopped and EGFP fluorescence was monitored as described in panel A. The specific digestion rate was 37.2 ± 5.2 ng EGFP min^{-1} ng activated pepsinogen^{-1}.

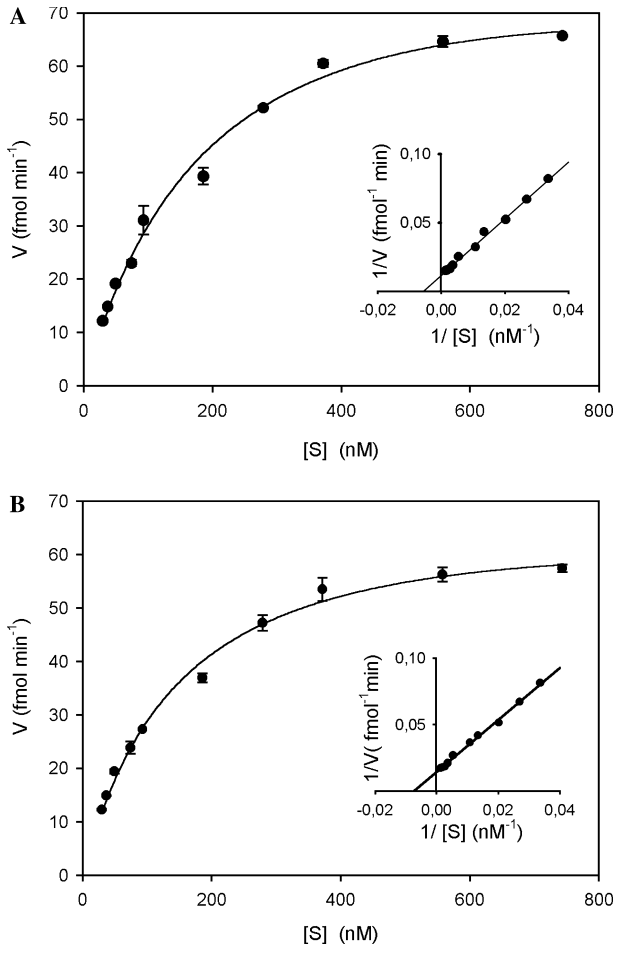


Fig. 4. Proteolytic activity depending on the substrate concentration. Acid-denatured EGFP was incubated with 4.9 ng pepsin (A) or activated pepsinogen (B) at 20 °C for 15 min. EGFP proteolysis was monitored with excitation at 490 nm and emission at 508 nm. The total amount of EGFP digested was determined using an acid-denatured and renatured EGFP standard curve. Double-reciprocal plots to calculate K_m and V_{max} are shown in the insets.

Likewise, if chromogenic synthetic peptides are used as substrates, pepsin can be determined in the microgram range [12]. The commonly used hemoglobin assay detects roughly the same amount of pepsin (1–10 μg) if the release of hemoglobin peptides is monitored at 280 nm. The assay is more sensitive only when radiolabeled hemoglobin is used as a substrate [28]. Alternatively, for clinical purposes, serum pepsinogen A is determined by a radioimmunoassay (Biohit, Helsinki, Finland), with 1–20 ng per assay.

Next, to determine kinetic constants, EGFP concentrations were varied from 0.88 to 22 μg (29.7–743.2 nM) while keeping a fixed pepsin concentration of 4.9 ng (141 pM). Substrate-dependent protease activity showed typical Michaelis–Menten kinetics (Fig. 4A). The apparent V_{max} and K_m values were calculated from the corresponding double-reciprocal plots and are given in Table 1. Experiments done using activated pepsinogen showed similar reaction kinetics (Fig. 4B and Table 1).

Table 1
Kinetic values for proteolytic fragmentation of EGFP catalyzed by porcine pepsin and activated porcine pepsinogen

	K_m (nM)	V_{max} [fmol s^{-1}]
Pepsin	158 ± 15	1.29 ± 0.105
Activated pepsinogen[a]	131 ± 4	1.10 ± 0.047

Note. Proteolysis was analyzed at 20 °C (pH 2.0). Details of the assay conditions are described in Fig. 4. Data represent mean values from two independent sets of experiments in duplicate.

[a] Pepsinogen activation was done for 10 min at 37 °C (pH 2.0) as described in Materials and methods.

To visualize pepsin-catalyzed proteolysis of acid-denatured EGFP, the substrate was incubated with pepsin or activated pepsinogen for up to 30 min at 20 °C. At different times, the digestion products were neutralized, concentrated, and separated by SDS–PAGE. As is evi-

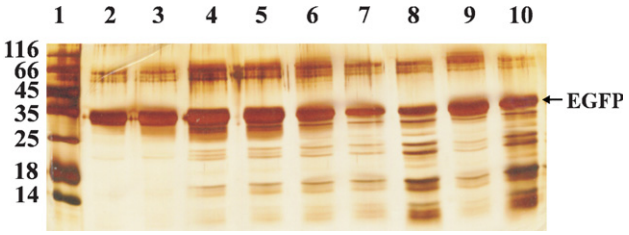


Fig. 5. SDS–PAGE analysis of pepsin-catalyzed EGFP cleavage. Acid-denatured EGFP (55 μg) was incubated with 24 ng pepsin or activated pepsinogen. At different times, proteolysis were stopped by neutralization, the probes were subjected to trichloroacetic acid precipitation, and an aliquot (1.1 μg) was separated on 17.5% SDS–PAGE. Lane 1, protein standard with the masses given in kilodaltons; lane 2, 1.1 μg native EGFP; lanes 3–8, EGFP after digestion by pepsin at 20 °C for 0, 2, 6, 10, 20, and 30 min; lanes 9 and 10, EGFP after digestion by activated pepsinogen at 20 °C for 10 and 30 min.

dent from Fig. 5, several distinct proteolytic fragments were produced. The same product pattern was observed after digestion with either pepsin (lanes 3–8 in the figure) or activated pepsinogen (lanes 9 and 10). An additional protein band (60 kDa) present in the undigested sample (lane 2) and in the digested probes (lanes 3–10) represents the dimeric form of EGFP regenerated after neutralization. After precipitation, it could not be completely dissociated by the SDS–PAGE buffer. During the course of proteolysis, it disappears to the same extent as does the EGFP monomer. As described in Materials and methods, homogeneity of the EGFP substrate was carefully checked by SDS–PAGE, reversed-phase HPLC, and mass spectrometry.

Because of its biological function, substrate specificity of pepsin is quite low. It preferentially hydrolyzes peptide bonds between hydrophobic aromatic amino acids. As calculated theoretically, the EGFP protein sequence contains more than 45 possible pepsin cleavage sites that would give a huge number of small peptides (1–8 aa) and only 9 peptides in a range of 1–3 kDa. The appearance of distinct proteolytic fragments (Fig. 5) indicates the presence of certain preferentially cleaved sites. Nearly the entire protein sequence is required to produce fluorescence [19]. Thus, a single pepsin cut might be sufficient to inhibit refolding and restoration of the fluorophore. Because of the presence of the different pepsin cleavage sites in EGFP with different affinities to the enzyme, the K_m value obtained can only be an average. Similarly, calculations of k_{cat} and K_m/k_{cat} values is not possible.

Monitoring pepsin by EGFP hydrolysis is a simple but quite selective procedure. Because of its very compact structure, GFP is very resistant to proteases at neutral pH, even at 1 mg ml^{-1} of peptidase concentration [29]. Thus, interference by common proteases that might be present in a sample cannot occur. In comparison with chromogenic or fluorogenic peptide substrates, EGFP is more stable. It can easily be purified and is convenient to handle and store. The EGFP-based pepsin assay described here is suited to detection of pepsin in highly diluted samples, even in the presence of a high background. In addition, it is sensitive enough to allow a high throughput application.

Acknowledgments

This work was supported by a grant from the Stiftung Industrieforschung (Germany) and by grants from the Bundesministerium für Bildung, Wissenschaft, Forschung, und Technologie (BMBF, Germany). The authors also thank A. Schierhorn (Halle, Germany) for mass spectrometry of EGFP.

References

[1] J Tang, Pepsin A, in: A.J. Barrett, N.D. Rawlings, J.F. Woessner (Eds.), Handbook of Proteolytic Enzymes, vol. 1, Elsevier, Amsterdam, 2004, pp. 19–28.
[2] J.S. Fruton, A history of pepsin and related enzymes, Q. Rev. Biol. 77 (2002) 127–147.
[3] A. Scorilas, E.P. Diamandis, M.A. Levesque, A. Papanastasiou-Diamandi, M.J. Khosravi, M. Giai, R. Ponzone, R. Roagna, P. Sismondi, C. Lopez-Otin, Immunoenzymatically determined pepsinogen C concentration in breast tumor cytosols: an independent favorable prognostic factor in node-positive patients, Clin. Cancer Res. 5 (1999) 1778–1785.
[4] J.H. Northrop, Crystalline pepsin: I. Isolation and tests of purity, J. Gen. Physiol. 13 (1930) 739–766.
[5] M.L. Anson, A.E. Mirsky, The estimation of pepsin with hemoglobin, J. Gen. Physiol. 16 (1932) 59–63.
[6] J.H. Northrop, Pepsin activity units and methods for determining peptic activity, J. Gen. Physiol. 16 (1932) 41–58.
[7] M.L. Anson, The estimation of pepsin, trypsin, papain, and cathepsin with hemoglobin, J. Gen. Physiol. 22 (1938) 65–78.
[8] P. McPhie, A turbidimetric milk-clotting assay for pepsin, Anal. Biochem. 73 (1976) 258–261.
[9] J.S. Fruton, Pepsin, in: P.D. Boyer (Ed.), The Enzymes, third ed., Academic Press, New York, 1971, pp. 119–164.
[10] G.P. Sachdev, J.S. Fruton, Secondary enzyme–substrate interactions and the specificity of pepsin, Biochemistry 9 (1970) 4465–4470.
[11] A. P. Ryle, Pepsins, gastricsins, and their zymogens, in: H.U. Bergmeyer (Ed.), Methods of Enzymology Analysis, third ed., vol. 5, 1984, pp. 223–238.
[12] B.M. Dunn, M. Jimenez, B.F. Parten, M.J. Valler, C.E. Rolph, J. Kay, A systematic series of synthetic chromophoric substrates for aspartic proteinases, Biochem. J. 237 (1986) 899–906.
[13] J. Pohl, B.M. Dunn, Secondary enzyme–substrate interactions: kinetic evidence for ionic interactions between substrate side chains and the pepsin active site, Biochemistry 27 (1988) 4827–4834.
[14] M.N. Raymond, E. Bricas, R. Salesse, J. Garnier, P. Garnot, B. Ribadeau Dumas, A proteolytic unit for chymosin (rennin) activity based on a reference synthetic peptide, J. Dairy Sci. 56 (1973) 419–422.
[15] R.A. da Silva Gomes, R.P. Batista, A. Costa de Almeida, D.N. da Fonseca, L. Juliano, V. Hial, A fluorimetric method for the determination of pepsin activity, Anal. Biochem. 316 (2003) 11–14.
[16] M. Zimmer, Green fluorescent protein (GFP): applications, structure, and related photophysical behavior, Chem. Rev. 102 (2002) 759–781.
[17] R.Y. Tsien, The green fluorescent protein, Annu. Rev. Biochem. 67 (1998) 509–544.

[18] B.P. Cormack, R.H. Valdivia, S. Falkow, FACS-optimized mutants of the green fluorescent protein (GFP), Gene 173 (1996) 33–38.

[19] X. Li, G. Zhang, N. Ngo, X. Zhao, S.R. Kain, C.C. Huang, Deletions of the *Aequorea victoria* green fluorescent protein define the minimal domain required for fluorescence, J. Biol. Chem. 272 (1997) 28545–28549.

[20] G.H. Patterson, S.M. Knobel, W.D. Sharif, S.R. Kain, D.W. Piston, Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy, Biophys. J. 73 (1997) 2782–2790.

[21] W.W. Ward, Biochemical and physical properties of green fluorescent protein, in: M. Chalfie, S. Kain (Eds.), Green Fluorescent Protein: Properties, Applications, and Protocals, Wiley–Liss, New York, 1998, pp. 45–75.

[22] C.N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, How to measure and predict the molar absorption coefficient of a protein, Protein Sci. 4 (1995) 2411–2423.

[23] W.W. Ward, S.H. Bokman, Reversible denaturation of *Aequorea* green-fluorescent protein: physical separation and characterization of the renatured protein, Biochemistry 21 (1982) 4535–4540.

[24] A.B. Cubitt, L.A. Woollenweber, R. Heim, Understanding structure–function relationships in the *Aequorea victoria* green fluorescent protein, Meth. Cell Biol. 58 (1999) 19–30.

[25] S. Enoki, K. Saeki, K. Maki, K. Kuwajima, Acid denaturation and refolding of green fluorescent protein, Biochemistry 43 (2004) 14238–14248.

[26] T. Kageyama, Pepsinogens, progastricsins, and prochymosins: structure, function, evolution, and development, Cell. Mol. Life Sci. 59 (2002) 288–306.

[27] J. Marciniszyn Jr., J.S. Huang, J.A. Hartsuck, J. Tang, Mechanism of intramolecular activation of pepsinogen: evidence for an intermediate delta and the involvement of the active site of pepsin in the intramolecular activation of pepsinogen, J. Biol. Chem. 251 (1976) 7095–7102.

[28] X.L. Lin, R.N. Wong, J. Tang, Synthesis, purification, and active site mutagenesis of recombinant porcine pepsinogen, J. Biol. Chem. 264 (1989) 4482–4489.

[29] S.H. Bokman, W.W. Ward, Renaturation of *Aequorea* green-fluorescent protein, Biochem. Biophys. Res. Commun. 101 (1981) 1372–1380.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Protein Expression and Purification 47 (2006) 662–671
Protein Expression & Purification
www.elsevier.com/locate/yprep

A novel fusion protein system for the production of native human pepsinogen in the bacterial periplasm

Ajamaluddin Malik, Rainer Rudolph, Brigitte Söhling *

Institut für Biotechnologie, Martin-Luther-Universität Halle-Wittenberg, Kurt-Mothes-Str. 3, 06120 Halle, Germany

Received 11 November 2005, and in revised form 20 February 2006
Available online 20 March 2006

Abstract

Human pepsinogen is the secreted inactive precursor of pepsin. Under the acidic conditions present in the stomach it is autocatalytically cleaved into the active protease. Pepsinogen contains three consecutive disulfides, and was used here as a model protein to investigate the production of aspartic proteases in the *Escherichia coli* periplasm. Various N-terminal translocation signals were applied and several different expression vectors were tested. After fusion to pelB, dsbA or ompT signal peptides no recombinant product could be obtained in the periplasm using the T7 promoter. As a new approach, human pepsinogen was fused to *E. coli* ecotin (*E. coli* trypsin inhibitor), which is a periplasmic homodimeric protein of 142 amino acids per monomer containing one disulfide bridge. The fusion protein was expressed in pTrc99a. After induction, the ecotin-pepsinogen fusion protein was translocated into the periplasm and the ecotin signal peptide was cleaved. Upon acid treatment, the fusion protein was converted into pepsin, indicating that pepsinogen was produced in its native form. In shake flasks experiments, the amount of active fusion protein present in the periplasm was 100 μg per litre OD 1, corresponding to 70 μg pepsinogen. After large scale cultivation, the fusion protein was isolated from the periplasmic extract. It was purified to homogeneity with a yield of 20%. The purified protein was native. Acid-induced activation of the fusion protein proceeded very fast. As soon as pepsin was present, the ecotin part of the fusion protein was rapidly digested, followed by a further activation of pepsinogen.

Keywords: Pepsin; Pepsinogen; Aspartic; Protease; Ecotin; Periplasm; Disulfide; Folding

The production of human proteins for diagnostic and therapeutic applications is a still increasing field of research, and the first true industrial use of recombinant DNA technology [1]. *Escherichia coli* as the best characterised and most widely used bacterial host can be combined with a lot of different vectors and expression systems [2–4]. Most proteins applied in therapy and diagnosis (e.g., hormones, growth factors, or recombinant antibodies) are secreted proteins [5], and most of them contain essential disulfide bonds that are required for structure formation in the extracellular oxidizing environment. High-level cytoplasmic expression of these proteins in *E. coli* is frequently accompanied by the generation of inclusion bodies. This can be a benefit for the matter of easy purification in some cases, but in other situations refolding of these proteins and production of a correct disulfide pattern can be quite difficult.

Secretion of these proteins into the *E. coli* periplasm gives a better chance of proper folding due to the more oxidizing conditions in this extracellular compartment. Besides the formation of correct disulfide bonds, production in the periplasm can also ease the purification. A prokaryotic signal peptide is required for translocation of the proteins across the cytoplasmic membrane, but the presence of this signal sequence does not always ensure efficient protein translocation. In these cases, linking of the target molecule to a larger fusion partner can be a solution, and may lead to efficient secretion and an increase in correct folding.

Human pepsinogen is a secreted protein from the stomach that contains three consecutive disulfide bridges. It is produced by the gastric chief cells of fundic mucosa,

* Corresponding author. Fax: +49 345 5527013.
E-mail address: brigitte.soehling@biochemtech.uni-halle.de (B. Söhling).

1046-5928/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2006.02.018

where it is autocatalytically converted into pepsin under acidic conditions [6,7]. Pepsin is a member of the large family of aspartic proteases [8]. These enzymes are of great medical and pharmaceutical interest, because some of them play important roles in the development of various human diseases, such as Alzheimer's disease [9,10], hypertension [11], HIV viral maturation [12], gastrointestinal disorders, and as a prognostic tool for breast tumor invasiveness [13]. Lysosomal and endosomal aspartic proteases are required for general protein turnover and for presentation of antigens on the cell surface [14–16]. Most aspartic proteases form inclusion bodies when expressed in *E. coli*. In some cases, renaturation is very inefficient [17–19].

In the present paper, human pepsinogen was used as a model protein to investigate the production of aspartic proteases in the *E. coli* periplasm. To identify and quantify pepsinogen in periplasmic samples, a highly sensitive assay method is needed, since only minute amounts of the native recombinant protein can be expected to be present during small scale cultivation. In the neutral environment of the *E. coli* periplasm, pepsinogen is not converted into its catalytic form. But after acidification (pH 2.0) of the periplasmic extracts, pepsin activity should be present. We recently developed a new and highly sensitive pepsinogen and pepsin detection assay which allows detection of pg amounts of the active protease [20]. Acid-denatured enhanced green fluorescent protein (EGFP)[1] is used as a substrate. Under acidic conditions, it is cleaved by pepsin, or activated pepsinogen, respectively. Upon neutralisation, uncleaved EGFP molecules refold completely and can be quantified by fluorescence. The loss in fluorescence is directly proportional to the amount of protease present in a sample.

To produce human pepsinogen in the periplasm, various N-terminal translocation signals were evaluated and several different expression vectors were chosen, but no conditions were found to detect native recombinant product in the periplasm. As a new approach, we fused the N-terminus of human pepsinogen to the C-terminus of *E. coli* ecotin (*E. coli* trypsin inhibitor), to use this protein as a potential translocator and folding enhancer. Ecotin is a periplasmic homodimeric protein of 142 amino acids per monomer that functions as a potent inhibitor of a variety of serine proteases [21,22]. Ecotin is expressed continuously during stages of cell growth [21]. Each monomer is arranged in a single domain best described as a mostly antiparallel seven-stranded ß-barrel [23]. The functional dimer is created through a domain swap mechanism involving the C-terminal region of each monomer. It resembles a butterfly with two protease binding sites (Fig. 1A) [23]. Ecotin undergoes oxidative folding in the periplasm so that one disulfide bridge is formed, that links two of the four surface loops present in one monomer. Ecotin is exceedingly stable and has been engineered by phage display as a scaffold protein to develop new and highly specific protease inhibitors [24,25]. In this study, we utilised ecotin as a fusion tag to produce human pepsinogen in native form in the periplasmic space of *E. coli*.

Materials and methods

Enzymes were obtained from NEB or MBI Fermentas. Porcine pepsin, porcine pepsinogen, and pepstatin A were purchased from Sigma. Benzonase was obtained from Merck, lysozyme was from Roche. Ni–NTA resin was from Novagen, and Hi-TrapQ HP, Superdex 75, and Superdex 200 prepacked columns were from Amersham Biosciences. All other chemicals were reagent grade. Molecular mass standard used for SDS–PAGE was purchased from Peqlab, the gel filtration low and high molecular weight standards were from Amersham Biosciences. Fluorescence measurements were taken from Hitachi-F4500 fluorescence spectrophotometer. pHQPEX-30-5 encoding human pepsinogen A was kindly provided by R. Bolli, ZLB Bioplasma AG, Bern, Switzerland. Expression vectors pET20b and pET12a were from Novagen. *E. coli* BL21(DE3) Gold was obtained from Stratagene.

Construction of expression plasmids

Human pepsinogen A was amplified from pHQPEX-30-5 using primers *Nco*I-Peps-Fwd (5′-ACG CCC ATG GGT TAC AAG GTC CCC CTC CAT-3′) and *Sal*I-Peps-Rev (5′-ACG CGT CGA CTA TTA AGC CAC GGG GGC-3′). The fragment was cloned into pCR-Blunt II-TOPO (Invitrogen) according to the manufacturer's instructions. The product was cut with *Nco*I and *Sal*I and ligated into pET20b to give pET-pelB-peps.

A synthetic nucleotide sequence encoding the signal sequence of *dsbA* (5′-CT AGA AAT AAT TTT GTT TAA CTT TAA GAA GGA GAT ATA CAT ATG AAA AAG ATT TGG CTG GCG CTG GCT GGT TTA GTT TTA GCG TTT AGC GCA TCG GCG GCC ATG-3′) was inserted into *Xba*I-*Nco*I cut pET20b, thereby exchanging the signal sequence of *pelB* by that from *dsbA*. Insertion of pepsinogen was carried out with primers *Nco*I-Peps-Fwd and *Sal*I-Peps-Rev as described above, to give pET-dsbA-peps. For a fusion to the *ompT* signal sequence, pepsinogen was amplified from pHQPEX-30-5 with primers *Sal*I-Peps-Fwd (5′-ACG CGT CGA CGA TGG GTT ACA AGG TCC CCC TCC AT-3′) and *Sal*I-Peps-Rev. The fragment was cut with *Sal*I and ligated into *Sal*I-cut pET12a to give pET-ompT-peps.

To clone pelB-pepsinogen into pTrc99a and pBAD22, the sequence was amplified from pET-pelB-peps with primers *Eco*RI-peps-Fwd (5′-TTC TAA GAA TTC GAA GGA GAT ATA CAT AAT GAA ATA CCT GCT GCC GAC-3′) and Peps-His-Rev (5′-GTA TAT GTC GAC TTA GTG GTG GTG GTG GTG GTG AGC CAC GGG GGC CAG ACC-3′). The resulting fragment encoded

[1] *Abbreviation used:* EGFP, enhanced green fluorescent protein.

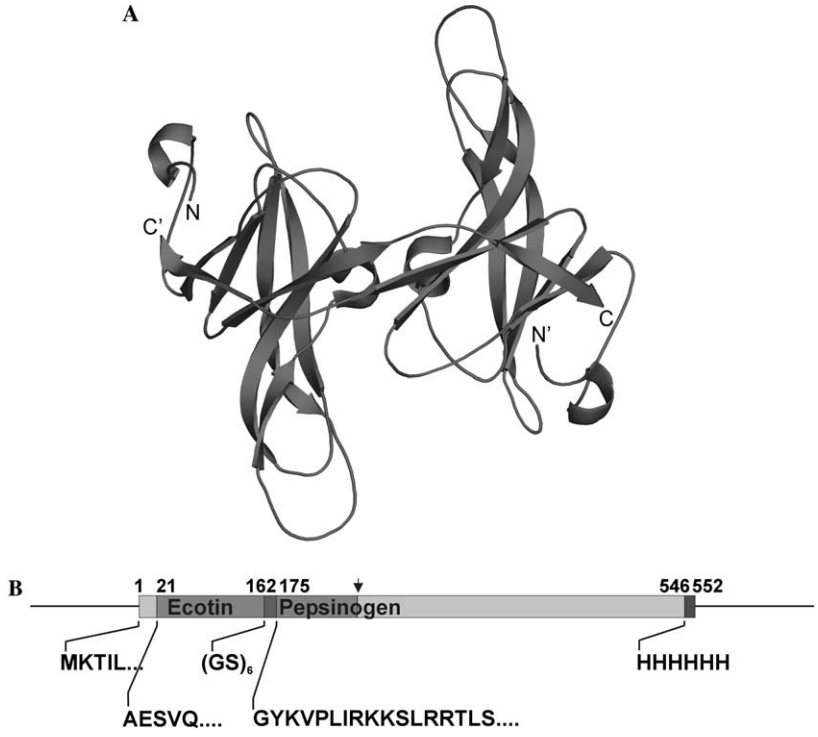


Fig. 1. Design of the ecotin-pepsinogen fusion protein. (A) Structural model of the ecotin dimer. The N- and C-termini of each subunit are indicated as N (N′) and C (C′). The C-terminal domains together form a two-stranded antiparallel ß-ribbon. The terminal residues 139–142 are positioned orthogonally with respect to the core sheet. They protrude to opposite sides of the dimer. The surface loops required for trypsin interaction direct to the top and the bottom of the figure. The structure was created from PDB entry 1ECZ using Pymol (http://pymol.sourceforge.net/). (B) Schematic draw of the ecotin-pepsinogen fusion protein as present in pEGP1. Amino acid numbering starts with the ecotin signal peptide (1–20), followed by mature ecotin (21–162) a GS6 linker (163–174), mature pepsinogen (175–546) and a C-terminal His-tag (547–552). The arrow indicates the location of the cleavage site for pepsinogen processing to pepsin.

pelB-pepsinogen followed by six histidine residues. It was cloned into pCR-Blunt II-TOPO, cut with *Eco*RI and *Sal*I and ligated into *Eco*RI-*Sal*I-cut pTrc99a to give pTrc-pelB-peps. The same fragment was ligated into *Eco*RI-*Sal*I-cut pBAD22 to give pBAD-pelB-peps.

For a fusion of pepsinogen to *E. coli* ecotin, the ecotin sequence (GenBank M60876) was amplified from the *E. coli* JM83 using primers Ecotin-Fwd (5′-TTC TAA GAA TTC GAA GGA GAT ATA CAT AAT GAA GAC CAT TCT ACC TGC A-3′) and Ecotin-GS-Rev (5′-ATC TTA TCC GGA ACC AGA ACC AGA ACC ACG AAC TAC CGC ATT GTC AAT TT-3′) and the fragment was inserted into pCR-Blunt II-TOPO. Human pepsinogen was amplified from pHQPEX-30-5 with primers GS-Peps-Fwd (5′-GTT ATA TCC GGA TCT GGT TCT GGT TCT GGT TAC AAG GTC CCC CTC A-3′) and Peps-His-Rev and was cloned into pCR-Blunt II-TOPO. This plasmid was transformed into *E. coli* GM 2163 (NEB) to stop methylation of GmATC at the *Bsp*E1 site. The ecotin fragment was cut with *Eco*RI and *Bsp*E1, and the pepsinogen fragment was cut with *Bsp*E1 and *Sal*I. Both fragments were ligated into *Eco*RI-*Sal*1-cut pTrc99a to give pTrc-eco-peps (further on abbreviated as pEGP1). Both fragments were ligated into *Eco*RI-*Sal*1-cut pBAD22 to give pBAD-eco-peps. All plasmids were confirmed by sequencing.

Small scale fusion protein production

The expression plasmids were transformed into *E. coli* BL21 (DE3) Gold, and an overnight culture was inoculated and grown at 37 °C in LB medium containing ampicillin (100 mg L^{-1}) at 140 rpm. In all experiments, a culture harboring the respective expression vector (pET20b, pET12b, pTrc99a, or pBAD22) served as a control. For protein production, the cultures were inoculated with 1% from the overnight culture and grown at 37 °C. Induction of recombinant protein synthesis was started at OD_{600} of 0.5–0.7 by adding 1 mM IPTG. For pBAD expression plasmids, 0.02% arabinose was used instead. Incubation temperature during protein expression was 24 °C. In the experiments using pET plasmids, cultures were continuously grown at 24 °C.

The periplasmic fraction was prepared by the osmotic shock method according to [26]. Cells were harvested at 13,000 rpm for 2 min. 100 µl of resuspension buffer (25% sucrose, 0.3 M Tris–HCl, 0.5 mM $MgCl_2$, and 1 mM EDTA, pH 8.0) was added for each ml culture OD 1.0. After 10 min incubation at room temperature, the suspension was centrifuged for 10 min at 5000 rpm. The cells were resuspended in the same volume of ice-cold osmotic shock buffer (0.5 mM $MgCl_2$, 1 mM EDTA, and 10 mM Tris–HCl, pH 8.0) and held on ice for 10 min. The super-

natant after 10 min centrifugation at 13,000 rpm and 4 °C represented the periplasmic fraction.

Large scale biomass production

Escherichia coli BL21(DE3) Gold harboring pEGP1 was used to inoculate 20 ml LB containing ampicillin (100 mg L^{-1}). After 12 h shaking at 37 °C, 2.5 ml were used to inoculate 250 ml of the same medium and the culture was further grown in a shake flask at 37 °C for 8 h. The culture was then transferred into 8 L of yeast enriched medium (0.5% glucose, 5% yeast extract, 1.1% KH_2PO_4, 0.05% NH_4Cl, and 0.068% $MgCl_2$) in a 10 L-Fermentor (C10-3K, B. Braun, Melsungen). Fermentation was done at 37 °C at a pH value of 7.0. Feeding and protein production were started simultaneously after complete glucose depletion at OD_{600} of 18.7. IPTG was added to give a final concentration of 1 mM, and temperature was lowered to 24 °C. Feeding solution was 30% yeast extract and 25% glycerol. The feeding rate was continuously increased from 1.5 to 4.5 ml min^{-1}. After 5 h (OD_{600}45) cells were harvested at 6000 rpm for 15 min. The biomass was stored in aliquots at −80 °C.

Large scale periplasm preparation

Large scale periplasm preparation was done according to the method of [27] with some modifications. 30 g biomass was resuspended in 30 ml osmotic shock buffer (50% sucrose, 20 mM Tris–HCl, 10 mM EDTA, pH 9.0) supplemented with 1 mM GSH and 1 mM GSSG. The suspension was stirred in the cold for 2 h (500 rpm, 4 °C). Next, 310 ml of ice-cold distilled water containing 1 mM GSH and 1 mM GSSG were added followed by further stirring for 2 h. After centrifugation at 10,000*g* for 15 min, 6 mM $MgCl_2$ and 6 μl (500 U/μl) benzonase were added to the supernatant and the solution was stirred for 45 min at room temperature, followed by dialysis against 50 mM Tris–HCl, 300 mM NaCl, pH 8.0, at 4 °C. After addition of 10 mM imidazole, the solution was centrifuged at 10,000*g* for 20 min, and the supernatant was filtered through a 0.45 μM membrane. The final volume of periplasm preparation was 360 ml.

Metal chelating chromatography

Forty millilitres His-bind resin (Novagen) equilibrated in 50 mM Tris–HCl, 300 mM NaCl, 10 mM imidazole, pH 8.0, were added to 360 ml periplasm preparation, and the suspension was stirred at 100 rpm at 4 °C for 4 h. After sedimentation, the supernatant was gently removed and the beads washed with three volumes of equilibration buffer (batch mode). Elution was performed in 4 steps, each by adding 45 ml of 50 mM Tris–HCl, 300 mM NaCl, 250 mM imidazole, pH 8.0. The eluted fractions were pooled and dialysed against 50 mM Tris–HCl, pH 7.5, followed by centrifugation at 20,000 rpm for 20 min. The supernatant was passed through a 0.2 μM membrane.

Ion exchange chromatography

Fifty millilitres of the dialysed Ni–NTA pool was applied to a 5 ml HiTrapQ HP (Amersham) equilibrated with 50 mM Tris–HCl, pH 7.5, at a flow rate of 1 ml min^{-1}. The column was washed with 70 ml of the same buffer, and the elution was performed by a two-step gradient in the same buffer (45 ml of 0–30% NaCl at 1.5 ml min^{-1}, and 5 ml of 30–100% at 1 ml min^{-1}). After equilibration the column was used in the same way to purify the remaining Ni–NTA purified sample.

Gel exclusion chromatography

For purification of the fusion protein, a Superdex 75 column was equilibrated with 0.05 M Tris–HCl, 0.3 M NaCl, pH 7.5. A volume of 2.5 ml protein pooled from HiTrapQ was loaded at a flow rate of 0.5 ml min^{-1}. Elution was performed with 0.05 M Tris–HCl, 0.3 M NaCl, pH 7.5, at a flow rate of 2 ml min^{-1}.

Analytical gel filtration was run in 0.05 M Tris–HCl, 0.3 M NaCl, pH 8.0, on a Superdex 200 column (Amersham Biosciences) with a flow rate of 1 ml min^{-1}. The purified protein (0.75 mg) was loaded onto the column and the elution was recorded at 280 nm. To calibrate the column, five proteins (0.4 mg of each) from the Gel Filtration HMW and LMW calibration kit (Amersham Biosciences) were mixed in a total volume of 2.5 ml. The calibration proteins and apparent molecular masses used for calibration are indicated in the Fig. 4 legend.

Reversed phase HPLC

Homogeneity of the purified protein was analysed by reversed phase HPLC on a C4 column (Vydac). The column was equilibrated with distilled water containing 0.05% trifluoroacetic acid. 0.93 μg fusion protein and 5 μM pepstatin were loaded onto the column at 1 ml min^{-1}, and the column was washed with 3 ml equilibration solution. The fusion protein was eluted with a 35 ml gradient of 0–100% acetonitrile, 0.05% trifluoroacetic acid. The eluted protein was analysed by mass spectroscopy (REFLEX II; Bruker Daltonik GmbH, Bremen, Germany).

N-terminal sequencing

N-terminal amino acid sequence determination was carried out after blotting onto a polyvinylidene difluoride membrane by automated Edman degradation.

Acid-induced activation of the ecotin-pepsinogen fusion protein

The purified ecotin-pepsinogen fusion protein was diluted in 10 mM Tris–HCl, pH 7.5, to give 30 μg ml^{-1} and the temperature was adjusted to 37 °C. 0.1 vol of 1 M citric acid, pH 2.0, was added and the mixture was incubated at

37 °C. After 0.5, 1, 2, 4, 6, 8, 10, and 16 min incubation, 24 μl were removed and immediately mixed on ice with 16 μl of 1 M Tris–HCl, pH 8.5. As a control, 21.6 μl of the diluted non-acidified fusion protein were mixed directly with 16 μl of 1 M Tris–HCl, pH 8.5, followed by addition of 2.4 μl of 1 M citric buffer, pH 2.0. The samples were loaded on a 12% SDS–PAGE and analysed by silver staining.

Determination of pepsin activity

Determination of pepsin activity was done using EGFP as a substrate [20]. Activation of the samples was performed after temperature adjustment to 37 °C by adding 0.1 vol of 1 M citric acid, pH 2.0 followed by 10 min incubation. All protein samples obtained during the purification procedure were diluted 5-fold in 10 mM Tris–HCl, pH 7.5. If necessary, acid-activated samples were diluted further in 0.1 M citric acid, pH 2.0.

Five hundred microlitres containing 11 μg EGFP in 10 mM Tris–HCl, pH 7.5, were denatured by adding 0.1 vol of 1 M citric acid, pH 2. Next, 50 μl acidified protease samples or an acidified pepsinogen standard (0–50 ng) were added, and the samples were incubated at 20 °C for 10 min. Renaturation of EGFP was achieved by adding 400 μl of 1 M Tris–HCl, pH 8.5. The final pH of the solution was pH 8.0. Fluorescence of renatured EGFP was determined using a Hitachi F-4500 fluorescence spectrophotometer with excitation at 490 nm. Emission was monitored at 508 nm. Untreated EGFP in the same concentrations served as a control. The amount of active protease in the samples was calculated using acidified porcine pepsinogen as a standard as described [20].

Protein quantification

Total protein was determined according to Bradford [28]. The amount of fusion protein present in a sample was calculated according to the pepsin activity present, with 1 mg pepsin (34,642 Da) corresponding to 1.16 mg pepsinogen (40,133 Da) or 1.67 mg of the ecotin-pepsinogen fusion protein (57,902 Da). Purity of proteins was judged by SDS–PAGE and subsequent quantification using Phoretix 1D quantifier.

Results and discussion

Periplasmic production of pepsinogen

For periplasmic production of human pepsinogen, different expression variants were constructed. First, the sequence encoding mature human pepsinogen was fused to three different signal peptides (pelB, dsbA or ompT), and these fusions were cloned into pET expression plasmids (Novagen) with *E. coli* BL21(DE3) Gold as an expression host. Different growth temperatures and inducer concentrations were tested, but no active pepsin could be detected in the periplasmic extracts. SDS–PAGE analysis of the total cell lysates revealed significant expression of a protein at the expected size, but N-terminal sequencing of the pelB fusion revealed that the signal sequence was not cleaved off, indicating that the synthesised protein remained in the cytosol.

Next, the promoter system was changed. The pelB-pepsinogen fusion was cloned into pTrc99a that contains the strong synthetic trc promoter [29]. After induction of protein synthesis and further incubation for 4 h at 24 °C, pepsin activity corresponding to about 16 μg fusion protein L^{-1} OD 1 was determined upon acidification of the periplasm. If pepstatin (5 μM) was added prior to acidification, no proteolysis occurred. Likewise, no proteolytic activity was present in cells harboring the vector pTrc99a. Pepstatin is a potent pepsin inhibitor, indicating that the pelB-pepsinogen fusion cloned into pTrc99a yielded native pepsinogen. However, production of pepsinogen completely failed in some but not all cultures. Even though different strains and a variety of expression conditions (media, temperature, induction time, and inducer concentration) were tested, no clear and reproducible data could finally be obtained. In a different approach, the pelB-pepsinogen fusion was cloned into pBAD22 under control of the araBAD promoter [30]. In this case, no significant pepsin activity was obtained, probably due to a very low expression level.

Design and expression of an ecotin-pepsinogen fusion protein

As a new concept to deliver a protein into the periplasm, a fusion to the *E. coli* ecotin gene was chosen. Since ecotin is expressed quite well [31] and folds easily, a fusion to ecotin might promote translocation and enhance folding of human pepsinogen.

The genes encoding ecotin and human pepsinogen A were amplified and fused to each other as depicted in Fig. 1B. The gene fusion encoded *E. coli* ecotin and its signal sequence, followed by human pepsinogen A and six histidine residues. Both proteins were linked by a stretch of six glycine-serine repeats. The gene fusion was cloned into pTrc99a, and pBAD22, respectively, to give pTrc-eco-peps (pEGP1) and pBAD-eco-peps.

Production of the fusion protein was analysed as before. In BL21(DE3) pEGP1 pepsin activity corresponding to 100 μg of the ecotin-pepsinogen fusion protein was determined per L bacterial culture (OD 1) 3.5 h after induction at 24 °C. Again, pepsin activity was completely inhibited by addition of pepstatin (5 μM), and no respective proteolytic activity was present in extracts from cells harboring the vector alone. Synthesis of ecotin-pepsinogen using pEGP1 is controlled by the *trc* promoter, and was found to be highly reproducible, giving similar values upon manifold repetition. After expression of the same fusion under control of the arabinose promoter (pBAD22) the amount was much lower (8 μg fusion protein L^{-1} bacterial culture OD1). Thus, further experiments were done with pEGP1.

Table 1
Purification of the ecotin-pepsinogen fusion protein

Step	Volume (ml)	Total protein (mg)[a]	Fusion protein (mg)[c]	Yield (%)
Periplasm preparation	360	612.6	8.9	100.0
Ni–NTA pool, dialysed	190	47.0	4.8	54.8
HiTrapQ chromatography	9	11.5	4.1	46.8
Superdex 75 chromatography	22	5.2[b]	2.0	22.9

[a] Protein concentration determined by Bradford.
[b] Protein was determined by A_{280}, using a calculated $\varepsilon = 83{,}240\ \mathrm{mol}^{-1}\ \mathrm{cm}^{-1}$for the ecotin-pepsinogen fusion protein.
[c] Protein concentration determined by pepsin assay, with 1 mg pepsin corresponding to 1.67 mg of the fusion protein.

Production and purification of the ecotin-pepsinogen

Escherichia coli BL21(DE3) containing pEGP1 was grown in a fermentor and protein production was performed in a fed batch process at 24 °C for 5 h. Under these conditions, a final biomass of 605 g was obtained from 9.5 L (OD 45).

After large scale periplasmic extraction of 30 g cells, about 8.9 mg of native fusion protein were determined (Table 1). A protein band corresponding to the expressed fusion protein (57.9 kDa) was visible after SDS–PAGE (Fig. 2, lane 1). If the periplasmic extract was acidified, proteolysis occurred, indicating pepsin activity (Fig. 2, lane 9).

Ni–NTA beads were added to the periplasm preparation and the fusion protein was bound by its C-terminal $(His)_6$-tag. The fusion protein present in the periplasm became strongly enriched by Ni–NTA extraction. It was almost quantitatively removed from the periplasm by this means (Fig 2, lanes 3 and 4), and only residual amounts of the fusion protein were detected after a second batch extraction (lane 5). During subsequent dialysis, most of the Ni–NTA purified protein remained soluble (lane 6), but about one third of the protein aggregated and was obtained as insoluble material after centrifugation (lane 7). Further purification was done by anion exchange chromatography on a HiTrapQ column. The ecotin-pepsinogen fusion eluted at 0.4 M NaCl. It was about 73% pure as judged by SDS–PAGE (Fig 2, lane 8). Final purification was achieved by size exclusion chromatography on Superdex 75. Fractions of the purified protein eluted from the column are shown in Fig. 3A.

The purification is summarised in Table 1. Starting with 612 mg protein present in the periplasm, 47 mg remained after Ni–NTA extraction, and 11.5 mg were present after HiTrapQ. Final gel filtration yielded 5.2 mg of highly pure fusion protein, as determined by A_{280}. The amount of fusion protein determined according to the pepsin assay was considerably lower (2.0 mg). It might be explained by a lower specific activity of human pepsin in the EGFP assay, as compared to porcine pepsin that was used throughout as a standard.

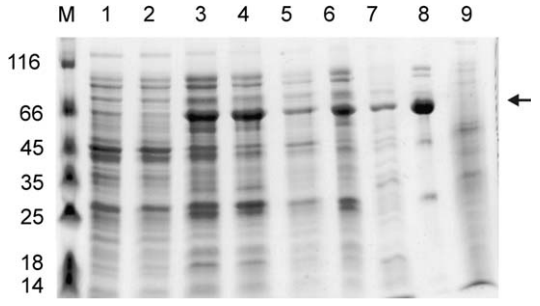


Fig. 2. Initial purification steps of the ecotin-pepsinogen fusion protein. The proteins were separated on a 12% SDS–PAGE. M, protein standard; lane 1, periplasmic extract after large scale expression; lane 2, periplasm after Ni–NTA extraction; lanes 3 and 4, Ni–NTA batch eluate, lane 5, Ni–NTA eluate during a 2nd batch, lanes 6 and 7, soluble and aggregated protein after dialysis, lane 8, protein pool after HiTrapQ chromatography, lane 9, periplasmic extract as shown in lane 1 but acid-treated (10 min, pH 2.0). The arrow indicates the fusion protein (57.9 kDa).

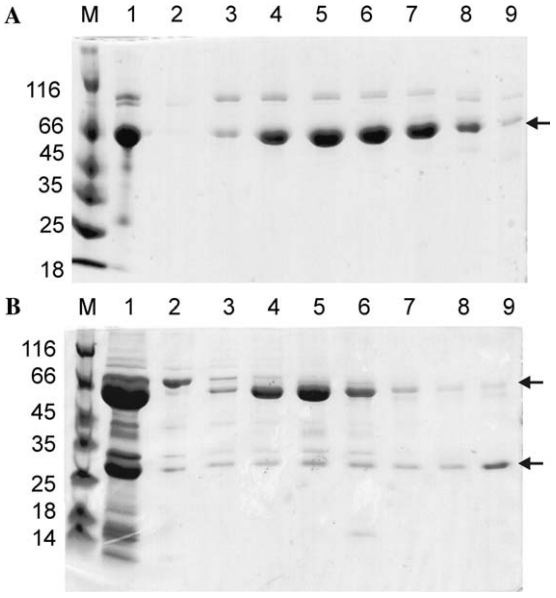


Fig. 3. Final purification step and effect of disulfide shuffling. (A) SDS–PAGE analysis of fractions eluted from Superdex 75 during purification of the ecotin-pepsinogen fusion protein. The purification was done as described in Materials and methods. Preparation of the periplasmic extract was done in the presence of 1 mM GSSG and GSH, respectively. The arrow indicates the ecotin-pepsinogen fusion protein (57.9 kDa). (B) SDS–PAGE analysis of fractions eluted from Superdex 75 using the same purification protocol, except that preparation of the periplasmic extract was done in the absence of thiols. The upper arrow (57.9 kDa) indicates the ecotin-pepsinogen fusion protein, the lower arrow marks SlyD (20 kDa), a cytoplasmic *E. coli* protein that was coeluted under these conditions.

Importance of disulfide shuffling

For the purification scheme described above, periplasmic extracts were prepared by osmotic shock, according to a protocol designed for large scale periplasm preparation using frozen cells [27]. Our modified protocol included the addition of a disulfide shuffling system (1 mM GSSG and GSH, respectively). In its absence, following the same protocol, an additional 20 kDa protein was present after the final gel filtration step (Fig. 3B). N-terminal sequencing identified it as SlyD, a cytoplasmic histidine-rich peptidyl-prolyl isomerase [32] that is often found as a contaminant protein during Ni–NTA chromatography [33]. SlyD contains 6 cysteine residues in the C-terminal part, and mixed protein disulfides of SlyD and ecotin-pepsinogen might be formed during protein extraction. This was not observed in the presence of a disulfide shuffling system, and the fusion protein was pure after the final gel filtration step (Fig. 3A).

Properties of the purified fusion protein

The N-terminal sequence of the purified fusion protein corresponded to that of the mature *E. coli* ecotin (Ala-Glu-Ser-Val-Gln-Pro-). The purified protein was further analysed by reversed phase HPLC. It was loaded onto the column in the presence of 5 μM pepstatin, to avoid activation to pepsin under the acidic experimental conditions. A single peak was obtained. A mass of 57,896 Da was determined by MALDI-TOF. The calculated value for the reduced fusion protein is 6 Da more (57,902 Da). The apparent molecular mass of the purified fusion protein was determined by gel filtration on a Superdex 200 column was run (Fig. 4). According to the elution of the calibration proteins, an apparent mass of 136,000 Da was calculated for the fusion protein, which most probably corresponds to a dimer (calculated mass 115,804 Da).

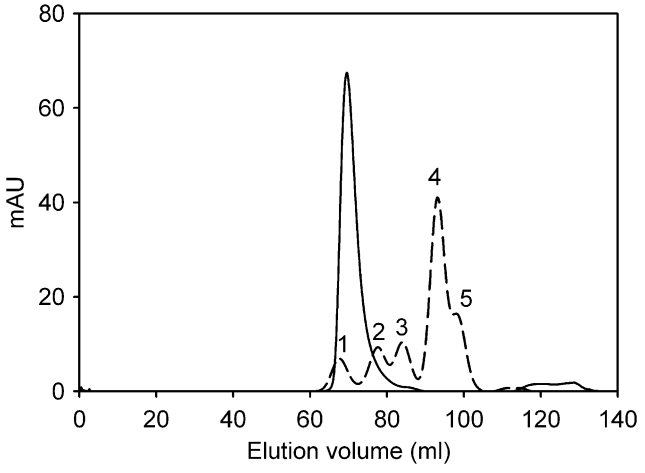


Fig. 4. Analytical gel filtration chromatography. The dashed line shows the elution profile of a mixture of five calibration proteins separated on a Superdex 200 column. These proteins were (1) aldolase (158 kDa); (2) bovine serum albumin (67 kDa); (3) ovalbumin (34 kDa); (4) chymotrypsinogen (25 kDa) and (5) ribonuclease A (13.7 kDa). The solid line represents the elution profile obtained the purified ecotin-pepsinogen fusion protein. The absorbance was recorded at 280 nm.

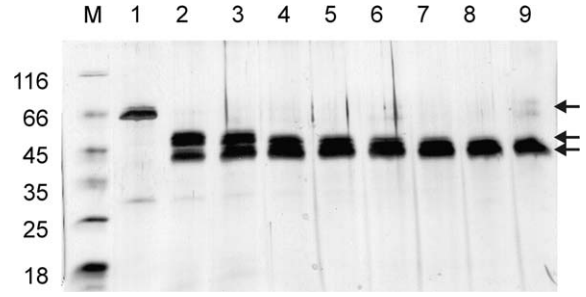


Fig. 5. Time-dependent activation of the ecotin-pepsinogen fusion protein. The cleavage reaction was analysed by SDS–PAGE (12%). Activation was done by addition of 1/10th volume of 1 M citric acid, pH 2.0, as described in Materials and methods. The reaction was stopped on ice by neutralisation (1 M Tris–HCl, pH 8.5). M, protein marker, lane 1, unacidified control, lanes 2–9, ecotin-pepsinogen acidified for 30 s, 1, 2, 4, 6, 8, 10, and 16 min, respectively. The arrows indicate the fusion protein (57.9 kDa), pepsinogen (40.1 kDa) and pepsin (34.6 kDa).

Ecotin-pepsinogen activation

The mechanism of acid-induced activation of ecotin-pepsinogen was further analysed by acidification of the purified fusion protein at 37 °C at different time intervals. After neutralisation, the probes were subjected to SDS–PAGE (Fig. 5). Activation proceeded very fast. The full-length fusion protein (58 kDa) disappeared within 30 s, and distinct fragments corresponding to pepsinogen (40 kDa) and pepsin (34 kDa) appeared instead.

The activation mechanism of pepsinogen occurs through both, bimolecular and unimolecular reaction mechanisms [34–36]. Structural studies of porcine pepsinogen at neutral pH show that the pepsin moiety of pepsinogen has nearly the same conformation as native pepsin, with the propeptide covering the active site cleft [37]. Under acidic conditions, the central helical part of the propeptide undergoes local conformational denaturation. It now binds to the active site as a substrate and is cleaved at Leu22-Lys23. Further hydrolysis at the peptide bond between the propeptide and the active enzyme seems to occur intermolecularly [38,39]. Ecotin is highly susceptible to pepsin cleavage [21]. Under the experimental conditions used here, the ecotin part of the proprotein is digested almost completely as soon as the first pepsin molecules are released, followed by hydrolysis of remaining part into pepsin within 8 minutes of acid treatment.

Concluding remarks

In summary, the ecotin fusion system proved to be an elegant method to efficiently produce and purify human pepsinogen in its native form, which could be easily proven by acid-induced activation and determination of pepsin activity. According to our purification scheme, 5.62 mg native fusion protein corresponding to 3.6 mg pepsinogen

were obtained from 30 g biomass. Thus, 7.6 mg native pepsinogen could be produced per L of fermentation broth. We did not optimize fermentation conditions, or further refine the periplasmic preparation step. Therefore, it should be possible to substantially increase the yield of purified product in the future.

If pepsinogen was directly fused to different signal sequences (ompT, dsbA, and pelB), no native product could be detected in the periplasmic extracts. This might be explained by the extremely high expression level of the T7 promoter system, so that the cell is unable to export the protein because of fast inclusion body formation in the cytoplasm. In eubacteria, most protein with cleavable signal sequences are thought to be exported by a relatively slow, partially posttranslational pathway [40]. Many of these, but not all, are bound by the chaperone SecB which prevents the large nascent polypeptide from folding into a conformation that would interfere with translocation [41]. Interestingly, it has recently been found that a dsbA signal peptide promotes the export of proteins via the SRP pathway [42]. Even thioredoxin that is located in the cytoplasm and folds very rapidly could be secreted into the periplasm after fusion to the *dsbA* signal peptide. Secretion of the periplasmic protein MalE is dependent on the chaperone SecB, but it was cotranslationally exported in a *secB* mutant strain with the help of a dsbA signal peptide [42]. If pepsinogen was fused to the dsbA signal peptide, it was not translocated, and no active pepsinogen was found in the periplasm.

The failure to obtain native pepsinogen in the periplasm regardless of the signal sequence used (pelB, dsbA, and ompT) could also be due to the highly basic N-terminal sequence of pepsinogen. It contains three arginine and three lysine residues within the first 14 amino acids. If fused to a signal peptide, this region defines a region of constraint properties that is critical to the secretability of proteins in *E. coli* because it participates in membrane spanning [43,44]. With respect to the positive charges, arginine is much more critical, because its guanidinium group cannot passively cross the membrane under physiological conditions [45].

Several different proteins have been used as fusion tags for periplasmic expression, with the maltose binding protein MalE as the most superior one [46]. MalE fusion proteins can be purified by affinity chromatography (New England Biolabs). MalE does not contain disulfides, and also confers solubility even during cytoplasmic expression [17]. MalE gets at least 80% translated before it becomes translocation competent [47]. It is bound by the chaperone SecB to keep it in an unfolded state. In *secB* deficient strains MalE precursors rapidly assume a folded and export-incompetent state [48]. Ecotin (16 kDa) is much smaller as compared to MalE (40 kDa), which might be an advantage for the expression of large target proteins. It is not known by which mechanism it is exported to the periplasm. The presence of a disulfide bridge and a highly structured fold suggests a SecB dependent posttranslational secretion, even if not a cotranslational SRP dependent translocation.

Another often used periplasmic fusion tag is the *staphylococcal* protein A (31 kDa). In Gram positive hosts, it is secreted into the media. Depending on the presence or absence of the signal peptide, protein A can also be used for periplasmic or cytoplasmic protein production in *E. coli*[49]. Smaller variants have been created that comprise only the signal peptide and two synthetic IgG-binding Z-domains (14 kDa). Respective fusion proteins are secreted and can be purified both, from the periplasm or from the medium [50–52].

The entire DsbA protein was also used efficiently as a fusion protein for periplasmic expression [53]. DsbA contains six disulfides and is the most important factor for disulfide bond formation in the periplasm. After fusion to the DsbA protein, the yield of native proinsulin secreted into the periplasm was strongly increased [54]. With proinsulin, aggregation is often observed, and was reported even after fusion to DsbA. Soluble proinsulin was produced by a fusion to the protein A (ZZ-proinsulin) [51]. It might be very interesting to see whether a fusion of proinsulin to the highly soluble ecotin might further increase the yield of native product. Recently, an interesting new fusion tag was described to secrete short cystine-knot peptides in native form. It consists of an enzymatically inactive variant of the *Bacillus amyloliquefaciens* RNase barnase [55]. Like ecotin, barnase is a small and highly stable protein (110 amino acids), but it does not contain a disulfide bridge.

The majority of aspartic proteases expressed so far in *E. coli* have been isolated as inclusion bodies, except for HIV protease. After fusion of pepsinogen to the maltose binding protein or thioredoxin (pepsinogen-mbp, pepsinogen-trx) and expression in the cytoplasm, the protein could not be obtained in its soluble form [17]. If thioredoxin was fused to the N-terminus of pepsinogen (trx-pepsinogen), native soluble pepsinogen could be obtained from the cytoplasm [56]. When porcine pepsinogen was expressed in *Pichia pastoris*, the protein was secreted into the medium as a native, non-glycosylated and a glycosylated variant. Both were active and had similar characteristics as the wild-type pepsinogen [57]. A yield of 30 mg per L culture medium was reported, which was more than that determined for cytosolic thioredoxin-fused protein (9 mg per 1.5 L medium).

This is the first report on a periplasmic expression system for aspartic proteases. After fusion to ecotin, pepsinogen was secreted, soluble and active, and could be easily purified and activated. The same system might be adopted to express other members of this class, such as cathepsin D, that cannot be refolded in sufficient amounts [18].

Acknowledgments

The present work was supported by a grant from the Stiftung Industrieforschung, Germany, and by grants from the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) of Germany. The authors

thank A. Schierhorn, Halle, for mass spectrometry, and P. Rücknagel, Halle, for N-terminal sequencing.

References

[1] G. Walsh, Pharmaceutical biotechnology products approved within the European Union, Eur. J. Pharm. Biopharm. 55 (2003) 3–10.
[2] F. Baneyx, Recombinant protein expression in *Escherichia coli*, Curr. Opin. Biotechnol. 10 (1999) 411–421.
[3] G. Hannig, S.C. Makrides, Strategies for optimizing heterologous protein expression in *Escherichia coli*, Trends Biotechnol. 16 (1998) 54–60.
[4] J. Swarts, Advances in *Escherichia coli* production of therapeutic proteins, Curr. Opin. Biotechnol. 12 (2001) 195–201.
[5] A.K. Pavlou, J.M. Reichert, Recombinant protein therapeutics—success rates, market trends and values to 2010, Nat. Biotechnol. 22 (2004) 1513–1519.
[6] J.S. Fruton, A history of pepsin and related enzymes, Q. Rev. Biol. 77 (2002) 127–147.
[7] J. Tang, Pepsin A, in: A.J. Barrett, N.D. Rawlings, J.F. Woessner (Eds.), Handbook of proteolytic enzymes, vol. 1, Elsevier, Amsterdam, 2004, pp. 19–28.
[8] N.D. Rawlings, A.J. Barrett, Introduction: aspartic peptidases and their clans, in: A.J. Barrett, N.D. Rawlings, J.F. Woessner (Eds.), second ed., Handbook of proteolytic enzymes, vol. 1, Elsevier, Amsterdam, 2004, pp. 3–12.
[9] S. Sinha, J.P. Anderson, R. Barbour, G.S. Basi, R. Caccavello, D. Davis, M.T. Doan, H.F. Dovey, N. Frigon, J. Hong, K. Jacobson-Croak, N. Jewett, P. Keim, J. Knops, I. Lieberburg, M. Power, H. Tan, G. Tatsuno, J. Tung, D. Schenk, P. Seubert, M. Suomensaari, S. Wang, D. Walker, J. Zhao, L. McConlogue, V. John, Purification and cloning of amyloid precursor protein ß-secretase from human brain, Nature 402 (1999) 537–540.
[10] R. Vassar, B.D. Bennett, S. Babu-Khan, S. Kahn, E.A. Mendiaz, P. Denis, D.B. Teplow, S. Ross, P. Amarante, R. Loeloff, Y. Lou, S. Fisher, J. Fuller, S. Edenson, J. Lile, M.A. Jarosinski, A.L. Biere, E. Curran, T. Burgess, J.C. Louis, F. Collins, J. Treanor, G. Rogers, M. Citron, Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE, Science 286 (1999) 735–741.
[11] F. Suzuki, K. Murakami, Y. Nakamura, T. Inagami, Renin, in: A.J. Barrett, N.D. Rawlings, J.F. Woessner (Eds.), second ed., Handbook of proteolytic enzymes, vol. 1, Elsevier, Amsterdam, 2004, pp. 54–61.
[12] S. Erickson-Viitanen, J. Manfredi, P. Viitanen, D.E. Tribe, R. Tritch, C.A. Hutchinson, D.D. Loeb, R. Swanstrom, Cleavage of HIV-1 gag polyprotein synthesized in vitro: sequential cleavage by the viral protease, AIDS Res. Hum. Retroviruses 5 (1989) 577–591.
[13] A. Scorilas, E.P. Diamandis, M.A. Levesque, A. Papanastasiou-Diamandi, M.J. Khosravi, M. Giai, R. Ponzone, R. Roagna, P. Sismoni, C. López-Otin, Immunoenzymatically determined pepsinogen C concentration in breast tumor cytosols: an independent favorable prognostic factor in node-positive patients, Clin. Cancer Res. 5 (1999) 1778–1785.
[14] G.E. Conner, Cathepsin D, in: A.J. Barrett, N.D. Rawlings, J.F. Woessner (Eds.), second ed., Handbook of proteolytic enzymes, vol. 1, Elsevier, Amsterdam, 2004, pp. 43–52.
[15] E.W. Hewitt, A. Treumann, N. Morrice, P.J. Tatnell, J. Kay, C. Watts, Natural processing sites for human cathepsin E and cathepsin D in tetanus toxin. Implications for T cell epitope generation, J. Immunol. 159 (1997) 4693–4699.
[16] M.A. Maric, M.D. Taylor, J.S. Blum, Endosomal aspartic proteases are required for invariant-chain processing, Proc. Natl. Acad. Sci. USA 91 (1994) 2171–2175.
[17] D. Sachdev, J.M. Chirgwin, Solubility of proteins isolated from inclusion bodies is enhanced by fusion to maltose-binding protein or thioredoxin, Protein Expr. Purif. 12 (1998) 122–132.
[18] P.E. Scarborough, B.M. Dunn, Redesign of the substrate specificity of human cathepsin D: the dominant role of position 287 in the S2 subsite, Protein Eng. 7 (1994) 495–502.
[19] X. Lin, Universal procedure for refolding of recombinant proteins, USP 6,585,268 (2003).
[20] A. Malik, R. Rudolph, B. Söhling, Use of enhanced green fluorescent protein EGFP to determine pepsin at high sensitivity, Anal. Biochem. 340 (2005) 252–258.
[21] C.H. Chung, H.E. Ives, S. Almeda, A.L. Goldberg, Purification from *Escherichia coli* of a periplasmic protein that is a potent inhibitor of pancreatic proteases, J. Biol. Chem. 258 (1983) 11032–11038.
[22] C.T. Eggers, I.A. Murray, V.A. Delmar, A.G. Day, C.S. Craik, The periplasmic serine protease inhibitor ecotin protects bacteria against neutrophil elastase, Biochem. J. 379 (2004) 107–118.
[23] M.E. McGrath, T. Erpel, C. Bystroff, R.J. Fletterick, Macromolecular chelation as an improved mechanism of protease inhibition: structure of the ecotin-trypsin complex, EMBO J. 13 (1994) 1502–1507.
[24] A.A. Stoop, C.S. Craik, Engineering of a macromolecular scaffold to develop specific protease inhibitors, Nature Biotechnol. 21 (2003) 1063–1068.
[25] C.-I. Wang, C. Yang, C.S. Craik, Isolation of a high affinity inhibitor of urokinase-type plasminogen activator by phage display of ecotin, J. Biol. Chem. 270 (1995) 12250–12256.
[26] Y. Kang, J.-W. Yoon, Effect of modification of connecting peptide of proinsulin on its export, J. Biotechnol. 36 (1994) 45–54.
[27] A.S. Rathore, R.E. Bilbery, D.E. Steinmeyer, Optimization of an osmotic shock procedure for isolation of a protein product expressed in *E. coli*, Biotechnol. Prog. 19 (2003) 1541–1546.
[28] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding, Anal. Biochem. 72 (1976) 248–254.
[29] E. Amann, B. Ochs, K.J. Abel, Tightly regulated tac promoter vectors useful for the expression of unfused and fused protein in *Escherichia coli*, Gene 69 (1988) 301–315.
[30] L.-M. Guzman, D. Belin, M.J. Carson, J. Beckwith, Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter, J. Bacteriol. 177 (1995) 4121–4130.
[31] H.R. Lee, J.H. Seol, O.M. Kim, C.S. Lee, S.W. Suh, Y.-M. Hong, K. Tanaka, A. Ichihara, D.B. Ha, C.H. Chung, Molecular cloning of the ecotin gene in *Escherichia coli*, FEBS Lett. 287 (1991) 53–56.
[32] S. Hottenrott, T. Schumann, A. Plückthun, G. Fischer, J.-U. Rahfeld, The *Escherichia coli* SlyD is a metal ion-regulated peptidyl-prolyl cis/trans-isomerase, J. Biol. Chem. 272 (1997) 15697–15701.
[33] C. Wülfing, J. Lombardero, A. Plückthun, An *Escherichia coli* protein consisting of a domain homologous to FK506-binding proteins (FKBP) and a new metal binding motif, J. Biol. Chem. 269 (1994) 2895–2901.
[34] M. Bustin, A. Conway-Jacobs, Intramolecular activation of porcine pepsinogen, J. Biol. Chem. 246 (1971) 615–620.
[35] X.L. Lin, R.N. Wong, J. Tang, Synthesis, purification, and active site mutagenesis of recombinant porcine pepsinogen, J. Biol. Chem. 264 (1989) 4482–4489.
[36] J. Marciniszyn Jr., J.S. Huang, J.A. Hartsuck, J. Tang, Mechanism of intramolecular activation of pepsinogen. Evidence for an intermediate delta and the involvement of the active site of pepsin in the intramolecular activation of pepsinogen, J. Biol. Chem. 251 (1976) 7095–7102.
[37] J.A. Hartsuck, G. Koelsch, S.J. Remington, The high-resolution crystal structure of porcine pepsinogen, Proteins 13 (1992) 1–25.
[38] C. Richter, T. Tanaka, R.Y. Yada, Mechanism of activation of the gastric aspartic proteinases: pepsinogen, progastricsin and prochymosin, Biochem. J. 335 (1998) 481–490.
[39] T. Kageyama, M. Ichinose, K. Miki, S.B. Athauda, M. Tanji, K. Takahashi, Difference of activation processes and structure of activation peptides in human pepsinogens A and progastricsin, J. Biochem. (Tokyo) 105 (1989) 15–22.
[40] L.G. Josefsson, L.L. Randall, Different exported proteins in *E. coli* show differences in the temporal mode of processing in vivo, Cell 25 (1981) 151–157.

[41] G. Liu, T.B. Topping, L.L. Randall, Physiological role during export fro the retardation of folding by the leader peptide of maltose-binding protein, Proc. Natl. Acad. Sci. USA 86 (1989) 9213–9217.
[42] C.F. Schierle, M. Berkmen, D. Huber, C. Kumamoto, D. Boyd, J. Beckwith, The DsbA signal sequence directes efficient, cotranslational export of passenger proteins to the *Escherichia coli* periplasm via the signal recognition particle pathway, J. Bacteriol. 185 (2003) 5706–5713.
[43] A.V. Kajava, S.N. Zolov, A.E. Kalinin, M.A. Nesmeyanova, The net charge of the first 18 residues of the mature sequence affects protein translocation across the cytoplasmic membrane of gram-negative bacteria, J. Bacteriol. 182 (2000) 2163–2169.
[44] R.G. Summers, J.R. Knowles, Illicit secretion of a cytoplasmic protein into the periplasm of *Escherichia coli* requires a signal peptide plus a portion of the cognate secreted protein, J. Biol. Chem. 264 (1989) 20074–20081.
[45] R.G. Summers, C.R. Harris, J.R. Knowles, A conservative amino acid substitution, arginine for lysine, abolishes export of a hybrid protein in *Escherichia coli*, J. Biol. Chem. 264 (1989) 20082–20086.
[46] C. di Guana, P. Lib, P.D. Riggsa, H. Inouyeb, Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein, Gene 67 (1988) 21–30.
[47] L.L. Randall, Translocation domains of nascent periplasmic proteins across the cytoplasmic membrane is independent of elongation, Cell 33 (1983) 231–240.
[48] D.N. Collier, V.A. Bankaitis, J.B. Weiss, P.J.J. Bassford, The antifolding activity of SecB promotes the export of the *E. coli* maltose-binding protein, Cell 53 (1988) 273–283.
[49] B. Nilsson, L. Abrahamsen, M. Uhlen, Immobilization and purification of enzymes with staphylococcal protein A gene fusion vectors, EMBO J. 4 (1985) 1075–1080.
[50] H. Hellebust, S. Bergseth, L. Örning, Expression of the second epidermal growth factor-like domain of human factor VII in *Escherichia coli*, J. Biotechnol. 66 (1998) 203–210.
[51] F.J.M. Mergulhao, M.A. Taipa, J.M.S. Cabral, G.A. Monteiro, Evaluation of bottlenecks in proinsulin secretion by *Escherichia coli*, J. Biotechnol. 109 (2004).
[52] T. Moks, L. Abrahamsen, E. Holmgren, M. Bilich, A. Olsson, M. Uhlen, G. Pohl, C. Sterky, H. Hultberg, S. Josephson, A. Holmgren, H. Jörnvall, B. Nilsson, Expression of human insulin-like growth factor I in bacteria: use of optimized gene fusion vectors to facilitate protein purification, Biochemistry 26 (1987) 5239–5244.
[53] L.A. Collins-Racie, J.M. Mc Colgan, K.L. Grant, E.A. Di Blasio-Smith, J.M. Mc Coy, E.M. La Vallie, Production of recombinant bovine enterokinase catalytic subunit in *Escherichia coli* using the novel secretory fusion partner DsbA, Biotechnology 13 (1995) 982–987.
[54] J. Winter, P. Neubauer, R. Glockshuber, R. Rudolph, Increased production of human proinsulin in the periplasmic space of *Escherichia coli* fusion to DsbA, J. Biotechnol. 84 (2000) 175–185.
[55] H.-U. Schmoldt, A. Wentzel, S. Becker, H. Kolmar, A fusion protein system for the recombinant production of short disulfide bond rich cystine knot peptides using barnase as a purification handle, Protein Expr. Purif. 39 (2005) 82–89.
[56] T. Tanaka, R.Y. Yada, Expression of soluble cloned porcine pepsinogen A in *Escherichia coli*, Biochem. J. 315 (1996) 443–446.
[57] M.A. Yoshimasu, J.-K. Ahn, T. Tanaka, R.Y. Yada, Soluble expression and purification of porcine pepsinogen from *Pichia pastoris*, Protein Expr. Purif. 25 (2002) 229–236.

Manuscript under revision-Protein Expression and Purification

Periplasmic production of native human proinsulin as a fusion to *E. coli* ecotin

Ajamaluddin Malik[1], Marco Jenzsch[2], Andreas Lübbert[2], Rainer Rudolph[1], Brigitte Söhling[1†]

[1] Institute for Biotechnology, Martin Luther University, Halle-Wittenberg, Kurt-Mothes-Str 3 D-06120 Halle (Saale), Germany.

[2] Centre of Bioprocess Engineering, Martin Luther University, Halle-Wittenberg, Weinbergweg 22, D-06120 Halle (Saale), Germany.

[†] Corresponding author: +49 345 5524856
E-mail address: brigitte.soehling@biochemtech.uni-halle.de

Abstract

Native proinsulin belongs to the class of the difficult-to-express proteins in *E. coli*. Problems mainly arise due to its small size, a high proteolytic decay, and the necessity to form a native disulfide pattern. In the present study, human proinsulin was produced in the periplasm of *E. coli* as a fusion to ecotin, which is a small periplasmic protein of 16 kDa encoded by the host, containing one disulfide bond. The fusion protein was secreted to the periplasm and native proinsulin was determined by ELISA. Cultivation parameters were studied in parallel batch mode fermentations using *E. coli* BL21(DE3)Gold as a host. After improvement of fed-batch high density fermentation conditions, 153 mg fusion protein corresponding to 51.5 mg native proinsulin were obtained per L. Proteins were extracted from the periplasm by osmotic shock treatment. The fusion protein was purified in one step by ecotin affinity chromatography on immobilized trypsinogen. After thrombin cleavage of the fusion protein, the products were separated by Ni-NTA chromatography. Proinsulin was quantified by ELISA and characterized by mass spectrometry. To evaluate the influence of periplasmic proteases, the amount of ecotin-proinsulin was determined in *E. coli* BL21(DE3)Gold and in a periplasmic protease deficient strain, *E. coli* SF120.

Keywords: Proinsulin, Insulin, Ecotin, Periplasm, Fusion tag, Folding

Abbreviations: ABTS, 2,2´-Azino-di-[3-ethylbenzthiazoline sulfonate(6)] diammonium salt; IPTG, Isopropyl-β-D-thiogalactopyranoside; MES, 2-(N-morpholine)-ethane sulfonic acid ; OD, Optical density; ppGpp, Guanosine Tetraphosphate; PMSF, Phenyl-Methyl-Sulfonyl-Fluoride

Introduction

Escherichia coli is the best characterized and most widely used bacterial host for the production of recombinant proteins [1, 2]. *E. coli* expression systems imply a rapid generation of biomass, low-cost culture conditions, and are very versatile, due the availability of an increasingly large number of cloning vectors and host strains. However, most recombinant proteins applied in therapy and diagnoses are secreted proteins with essential disulfide bonds, and will form inclusion bodies when expressed in the cytoplasm of *E. coli*. This can be a benefit for ease of purification, if *in vitro* refolding conditions lead to the formation of the correct disulfide bonds. Alternatively, the protein can be directed to the periplasm and will be natively folded due to the oxidizing conditions in this extracellular compartment. A major drawback is the limited secretion capacity of the host, and the limited space of this compartment.

Naturally *E. coli* does not secrete high amounts of proteins [3] and the transport to the periplasm or to the culture medium is a particularly complex process [4, 5]. Translocation of a protein across the cytoplasmic membrane requires a signal peptide, but the presence of signal sequence alone does not always guarantee efficient protein translocation [6, 7]. An alternative is the use of larger fusion partner which may lead to efficient translocation and increase the chances of correct folding.

We recently discovered a novel fusion protein system for the production of disulfide-containing proteins in the periplasm of *E. coli* [8]. The N-terminal domain consists of *E. coli* ecotin including its signal sequence. Ecotin (*E. coli* trypsin inhibitor) is a dimeric, relatively small (16 kDa), broad range serine protease inhibitor. It contains one disulfide bond and is exceedingly stable [9]. In a previous study, human pepsinogen was fused to different signal peptides (pelB, dsbA, or ompT), but no recombinant product could be obtained in the periplasm. However, after fusion of pepsinogen to *E. coli* ecotin, native protein was found in the periplasm, and the fusion protein was purified to homogeneity with a yield of 23 % [8].

In the current study, we used this new protein fusion tool for the production of native proinsulin in the periplasm. Proinsulin is the precursor of insulin, that consists of two separate peptide chains which are the A-chain (21 amino acids) and the B-chain (30 amino acids) joined by 2 interchain disulfide bonds. A third intrachain disulfide bond is present in the A-

chain. In the β-cells of the human pancreas, the precursor of insulin is synthesized as a single peptide chain, preproinsulin. After cleavage of the N-terminal signal sequence, proinsulin is stored in secretary granules. Conversion of proinsulin to insulin proceeds before secretion into blood, and is catalyzed by thiol proteases which remove the connecting peptide (C-peptide) from B- and A-chains [10]. *In vitro*, the connecting peptide of proinsulin is cleaved with trypsin and carboxypeptidase B to obtain native insulin [11].

Insulin was the first biotechnology drug, isolated initially from cows and pigs, and introduced in 1922 for the treatment of diabetes. Production of recombinant human insulin in *E. coli* started in 1981 [12]. To date, insulin for therapeutic applications is manufactured in large-scale in *E. coli* and yeast [13]. The formation of the correct disulfide bonds in insulin only occurs at the level of proinsulin. Therefore, production of insulin in *E. coli* is mediated by formation of a single polypeptide chain comprising of proinsulin or a miniaturized derivative thereof [14]. Several strategies for production of proinsulin in the cytoplasm as insoluble inclusion bodies or soluble in the periplasm of *E. coli* have been described [15-19]. The production level of proinsulin as inclusion bodies is very high but refolding with correct disulfide bonds and purification is costly and complex.

Production of native proinsulin in *E. coli* is still a challenging task. Native proinsulin is a quite difficult to express protein, mainly due to severe proteolysis after production in the cytoplasm [20]. Upon secretion into the periplasm, the half-life of proinsulin increased from 2 to 20 minutes [20]. The oxidizing environment of the periplasm also leads to the formation of proper disulfide bonds in proinsulin. However, the yield is very low as compared to inclusion body production in the cytoplasm [21, 22].

An efficient secretory expression of proinsulin was reported by Kang and Yoon [14]. In this study, a modified proinsulin, shortened in the C-peptide, was fused to two Z-domains derived from the staphylococcal protein A, with significantly increased expression yields. A similar approach was recently published using unmodified proinsulin fused to none, a single Z-domain and a double Z domain, to evaluate potential bottlenecks in proinsulin secretion [18]. In another fusion approach, proinsulin was fused to full-length DsbA, including its signal sequence [19].

In the present study, proinsulin was secreted into the periplasm as a fusion to *E. coli* ecotin. As determined by ELISA, and proved by mass spectrometry, the fusion protein was native. The production of ecotin-proinsulin was studied in batch fermentations. After high-density fed-batch fermentation, the fusion protein was extracted from the periplasm by osmotic shock and was purified in one step by affinity chromatography on immobilized trypsinogen.

Materials and methods

Restriction Enzymes were obtained from NEB and MBI Fermentas. PCR purification and gel extraction kits were obtained from Qiagen. pCR-Blunt II-TOPO kit and Top10 competent cells were from Invitrogen. *E. coli* BL21(DE3)Gold was obtained from Stratagene and *E. coli* GM2163 (*dam*$^-$ and *dcm*$^-$) was from NEB. *E. coli* SF120 was kindly provided by Prof. George Georgiou's laboratory, Institute of Cellular and Molecular Biology, University of Texas, Austin, USA. Bovine trypsinogen and thrombin were purchased from Sigma. HisTrap FF crude and HiTrap NHS-activated HP columns were from Amersham Biosciences. Antibodies used for ELISA were a gift from Roche Molecular Biochemicals, Penzberg, Germany. Human proinsulin used as a standard was obtained from the NIBSC, National Institute for Biological Standards and Control, United Kingdom. Maxisorp® ELISA plates were purchased from Nunc, Copenhagen. Protein mass standard PeqGold used for SDS-PAGE was from Peqlab, the Mark12™ marker was from Invitrogen. NuPAGE (4-12%) Bis-Tris gels were purchased from Invitrogen. ZipTip$_{C18}$ was purchased from Millipore. Micro BCA ™ protein assay kit was purchased from Pierce. RP-HPLC Nucleosil C18 column was from Macherey-Nagel, Germany. All other chemicals were reagent grade.

Construction of the ecotin-proinsulin expression plasmid

pEGP1 containing the ecotin-pepsinogen fusion protein [8] was transformed into *E. coli* GM2163 (*dam*$^-$ and *dcm*$^-$). The plasmid was prepared and digested by BspE1 and Sal1, respectively, to remove the pepsinogen fragment. The digested plasmid encoded *E. coli* ecotin followed by three Gly-Ser repeats. It was dephosphorylated and purified. The proinsulin gene was amplified from pET20b-proinsulin [23]. The forward primer was 5'-GGT TCC GGA TCT GGT TCT GGT TCT CTG GTC CCC CGC GGT AGT CAC CAC CAC CAC CAC CAC CGT TTT GTG AAC CAA CAC CTG TGC GGC-3' to introduce three Gly-Ser repeats, a thrombin cleavage site, a hexa-histidine tag and an arginine residue just before the start codon of proinsulin. The reverse primer was 5'-AGT GTC GAC TTA GTT GCA GTA GTT CTC CAG CTG GTA-3'. The PCR product was cloned into pCR-Blunt II-TOPO. After sequencing, the plasmid was transformed into *E. coli* GM2163. This plasmid was digested with BspE1 and Sal1. The insert was purified, and ligated into the above mentioned digested vector. The resulting plasmid encoded a fusion of *E. coli* ecotin to human proinsulin, and was

named pEG-PI. A schematic draw of the ecotin-proinsulin construct is shown in Fig. 1. Sequencing was done in both directions.

Pilot experiment for the analysis of ecotin-proinsulin production

E. coli BL21(DE3)Gold containing pEG-PI was inoculated into 20 ml LB containing ampicillin (100 mg/L). After 10 h shaking at 37 °C, 0.1 ml culture was transferred into 500 ml medium and the culture was incubated at 37 °C for 8 h. It was then transferred into 8 L of yeast enriched medium (0.5% glucose, 5% yeast extract, 1.1% KH_2PO_4, 0.05% NH_4Cl, and 0.068% $MgCl_2$) in a 10 L fermenter (C10-3K, B. Braun, Melsungen, Germany). Fermentation was done at 37 °C at a pH value of 7.0. After depletion of glucose, feeding with 30% yeast extract and 25% glycerol was started at 1 ml/min and the temperature was lowered to 24 °C. Recombinant protein synthesis was started at an OD_{600} of 31 by adding 1 mM IPTG. Now, the feeding rate was continuously increased to 3 ml/min. After 4 h (OD_{600} = 59) the cells were harvested at 6000 rpm for 15 min. The biomass was stored at -80 °C. Periplasmic extract was prepared from the frozen biomass by osmotic shock according to Rathore *et al*. [24], modified as previously described [8]. Proinsulin was determined by ELISA. Fractions eluted from a Ni-NTA column were analyzed by SDS-PAGE. N-terminal amino acid sequencing of the 28 kDa fusion protein and a 20 kDa fragment was done after blotting onto a polyvinylidene difluoride membrane by automated Edman degradation.

Parallel small scale batch fermentations

To analyze the production of ecotin-proinsulin, *E. coli* BL21(DE3)Gold harboring pEG-PI was grown in a Sixfors® (Infors, Switzerland) parallel fermenter system consisting of six 400 ml vessels running in parallel. The vessels were inoculated with shake flask cultures previously grown for 10 h at 30 °C and 170 rpm. All cultivations were done in batch in M9 medium as described in [25]. Glucose was used as carbon source, with 25 g/L as initial substrate concentration. After 5 h incubation, the glucose concentration had decreased to 16 g/L, and the optical density had reached a value of 11 to 12. Now, fusion protein synthesis was induced by addition of IPTG. The cultures were further grown for 4 h. The final OD was 20-25, with a final glucose concentration of 5-6 g/L. From each vessel, 5 probes were withdrawn every hour. The periplasmic extract was prepared according to Winter *et al*. [19]. Proinsulin was analyzed by ELISA.

High density fed-batch fermentation

Large scale high density fed-batch fermentations were performed in a BBI Sartorius System's BIOSTAT® C 15-L- bioreactor with 8 L working volume. The fermenter was equipped with 3 standard 6-blade Rushton turbines that could be operated at up to 1400 rpm. The aeration rate could be increased up to 24 L/min. Aeration rate and then stirrer speed were increased one after the other in order to keep the dissolved oxygen concentration at 25% of its saturation value. To suppress foam formation and to increase oxygen solubility, head pressure was kept at 0.5 bar above the ambient pressure in the laboratory.

After an initial batch process in the presence of 25 g/L glucose, the fermentations were operated in the fed-batch mode beginning with an initial volume of 5 L at pH 7.0 and a temperature of 37 °C. All fermentations were started in the night by automated inoculation into the reactor. Substrate feeding started with a fixed exponential profile immediately after the concentration of glucose in the reactor reached values below 2 g/L. The feeding solution contained 40% glucose. The feeding profile was chosen to obtain a specific biomass growth rate of 0.5/h. When the density of the biomass reached a defined value, protein synthesis was started by addition of 1 mM IPTG. Then, the cultivation temperature was set to 25 °C and the feed rate profile was reduced to follow a specific biomass growth rate of 0.1/h.

The temperature was measured with a Pt-100, the pH was determined with an Ingold-pH-probe, pO_2 with an Ingold pO_2-Clark-electrode, CO_2 in the vent line with MAIHAK®'s Unor 610, and O_2 with MAIHAK®'s Oxor 610. Further, the total ammonia consumption during pH control was recorded by means of a balance beneath the base reservoir. All these quantities were measured online. Additionally, enhanced foam levels could be detected with a foam sensor and, if the critical level was reached, a silicone antifoaming emulsion (ROTH®) was added. Biomass concentrations were measured offline (OD_{600}) with a Shimadzu® photo-spectrometer (UV-2102PC). Glucose was determined enzymatically with an YSI 2700 Select Bioanalyzer.

Large scale periplasm preparation

Following high cell density fed-batch fermentation, the biomass was directly processed without freezing. The periplasmic extract was prepared by osmotic shock according to method

of Winter *et al*. [19]. It was prepared 10-fold concentrated, so that 1 L of periplasmic extract corresponded to 10 L of a biomass of OD1. To this end, 54.8 ml fermentation broth from high density culture (OD_{600} = 182) was centrifuged, and the biomass was suspended in 1 L of 25% sucrose, 0.3 M Tris-HCl, 0.5 mM $MgCl_2$, 1 mM EDTA, pH 8.0, at room temperature for 30 min, followed by centrifugation (5,000 rpm at 4 °C for 20 min). The supernatant was discarded and the pellet was dissolved in 1 L of ice-cold 0.5 mM $MgCl_2$, 1mM EDTA, 10 mM Tris-HCl, pH 8.0 for 30 minutes. Next, centrifugation was done at 7,000 rpm for 30 min at 4 °C. The supernatant represented the periplasmic fraction. It was filtered through 0.2 µm and was stored at -80°C. For the affinity purification of ecotin-proinsulin, 250 ml periplasmic extracts were used.

Quantification of proinsulin by ELISA

Native proinsulin was determined by a sandwich ELISA according to Winter *et al* [19]. A pair of monoclonal antibodies was used, that recognizes native insulin and native proinsulin. It was a gift from Roche, and was commercially available as part of the Enzymun-Test Insulin (1 577 80851, Boehringer Mannheim). Human proinsulin from the NIBSC served as a standard. For the ELISA, 130 µl the monoclonal anti-human insulin antibody MAK<Insulin>AE9D6-IgG(DE) was coated in the wells of a Maxisorp® microplate. After a washing step, 100 µl of a peroxidase-conjugated second monoclonal anti-human insulin antibody, MAK<Insulin>CC9C10-IgG(DE), was applied and 10 µl probes or different standard dilutions, respectively, were added. After washing, 100 µl ABTS (1mg/ml in 3.25 mM sodium perborate, 39.8 mM citric acid, 60 mM Na_2HPO_4, pH 4.4) was added as substrate. ELISA measurements were recorded using a Versamax microplate reader (Molecular Devices). All probes were measured in duplicate with 3-5 dilutions.

To calculate the amount of fusion protein present in a probe, it was assumed that the molar amount of proinsulin determined in the probe corresponded to the amount of dimeric fusion protein. Cleavage of the dimeric fusion protein produced monomeric proinsulin, which resulted in a duplication of the ELISA signal (see results).

SDS-PAGE analysis

Protein probes were analyzed on NuPAGE 4-12% Bis-Tris gels (Invitrogen). The MES running buffer contained 50 mM Tris Base, 50 mM MES, 1 mM EDTA, 0.1% SDS, pH 7.3. The SDS-PAGE sample buffer (5 x) contained 0.1 M Tris, 4.8% SDS, 16% glycerol, 0.1% Bromophenol blue and 1 M β-mercaptoethanol, pH 8.0. The probes were boiled for 3 min.

Affinity purification of ecotin-proinsulin via immobilized trypsinogen

A prepacked ready-to-use 5 ml HiTrap column containing NHS-activated Sepharose HP (Amersham) was coupled with 45 mg trypsinogen (Sigma) according to the manufacturer's instructions. 250 ml periplasmic extract obtained from the biomass after high cell density fermentation (see above) was passed through the column at 4 °C with a flow rate of 3-4 ml/min using a peristaltic pump. The column was washed with 100 mM Tris-HCl, pH 8.5, and elution of the bound protein was performed with 50 mM HCl by FPLC. Eluted fractions (0.9 ml) were collected into tubes containing 0.1 ml 1M Tris-HCl, pH 8.5 for immediate neutralization. Fractions were visualized by SDS-PAGE.

Thrombin cleavage of ecotin-proinsulin

The peak fractions obtained from the trypsinogen column were pooled and thrombin (0.1 mg) was added. The mixture was incubated for 2 hours at room temperature to cleave the fusion protein.

Separation of ecotin and proinsulin by Ni-NTA chromatography

NaCl and imidazole were added to the thrombin-digested ecotin-proinsulin to give final concentrations of 300 mM and 5 mM, respectively. The solution was loaded onto an equilibrated HisTrap FF 5 ml column at 4 °C by FPLC. The column was washed with 50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, pH 8.0. Elution was performed with a linear gradient of 0-300 mM imidazole in 30 min at a flow rate of 1 ml/min. 500 µl aliquots from the peaks were subjected to TCA precipitation and was analyzed on NuPAGE (4-12%) Bis-Tris gels.

Mass spectrometry

Proteins present after trypsinogen chromatography were analyzed before and after thrombin digestion (10 µl each). Likewise, the proteins present after Ni-NTA chromatography were investigated. All probes were passed through a $ZipTip_{C18}$ (Millipore) with 0.6 µl bed volume. Following the manufacturer's instruction, the probes were eluted with 4 µl 0.1% TFA in 50% acetonitrile. Mass spectrometry was done by ESI-MS (ESQUIRE-LC, Bruker Daltonik GmbH, Bremen, Germany).

Protein masses, extinction coefficients and Micro BCA™ protein assay

The masses of the full-length and digested proteins were calculated from http://www.expasy.ch/cgi-bin/protparam. Full-length oxidized ecotin-proinsulin has a mass of 27,921.7 Da, oxidized GS-H_6-R-proinsulin has a mass of 10,511.8 Da, and oxidized mature proinsulin has 9388.6 Da. The molar extinction coefficients were calculated according to [26], with 29,380 mol^{-1} cm^{-1} for the monomeric full-length oxidized ecotin-proinsulin. For oxidized mature proinsulin and for GS-H_6R-proinsulin the extinction coefficient was ε = 6,335 mol^{-1} cm^{-1}.

For protein determination with the Micro BCA™ protein assay kit, purified proinsulin was dialyzed in a Slide-A-Lyzer mini dialysis tube (Pierce) against 50 mM Na-Phosphate, 100 mM NaCl, pH 7.5.The protein was quantified following manufacturer's protocol. BSA was used as a standard.

Small scale expression of ecotin-proinsulin in E. coli Bl21(DE3)Gold and E. coli SF120

pEG-PI was introduced into *E. coli* BL21(DE3)Gold and *E. coli* SF120. The bacteria were selected on LB agar containing ampicillin for *E. coli* BL21(DE3), or ampicillin, kanamycin and chloramphenicol for *E. coli* SF120, respectively. Both strains were inoculated into LB or M9 medium, respectively, containing the appropriate antibiotics, and the cultures were grown overnight at 30 °C. The overnight cultures were used to inoculate 20 ml of fresh medium in duplicate. The inoculation volume depended on the density of the overnight culture, in order to obtain similar cell densities upon inoculation. The cultures were grown at 30 °C to attain mid log phase. Induction with 1 mM IPTG was done after 3 hours of incubation for LB

cultures, and after 3.5 hours for M9 cultures, respectively. The temperature was lowered to 24 °C. Cells were harvested after 3 hours. Periplasm was prepared according to method of Winter *et al* [19] and native proinsulin was quantified by ELISA.

Results and discussion

Cloning and characterization of ecotin-proinsulin

A fusion protein for the secretion and production of native human proinsulin in the periplasm of *E. coli* was designed and was cloned into pTrc99a as shown in Fig. 1. The gene encoded a protein comprising the *E. coli* ecotin signal sequence and mature domain in the N-terminal part, and human proinsulin in the C-terminal part. The region in between was designed to encode a Gly-Ser linker $(GS)_6$, a thrombin cleavage site (LVPRGS), a His-tag $(H)_6$, and an arginine residue (R). The mature fusion protein encompassed 253 amino acids with a calculated mass of 27,921.7 Da in its native oxidized state. The resulting plasmid was named pEG-PI.

The plasmid pEG-PI was introduced into *E. coli* BL21(DE3)Gold. As a pilot experiment, the cells were grown in a fermenter at 37 °C in rich medium. After addition of 1 mM IPTG at an optical density (OD) of 31, the cells were further incubated for 4 h at 24 °C. From the frozen biomass a periplasmic extract was prepared and was analyzed for the presence of native proinsulin by ELISA. The periplasmic extract contained proinsulin corresponding to 94.4 µg fusion protein produced per L fermentation broth. From the final OD_{600} of 59, this refers to a specific yield of only 1.6 µg of proinsulin produced per L OD1, which was considerably low. Furthermore, only an extremely low quantity of full-length native fusion protein remained, when the periplasmic extract was further processed by Ni-NTA chromatography or ion exchange chromatography. SDS-PAGE after Ni-NTA chromatography revealed that the majority of the 28 kDa fusion protein was cleaved. A 20 kDa proteolytic fragment was present having the same N-terminal amino acid sequence as the 28 kDa fusion protein. Both revealed the N-terminus of mature *E. coli* ecotin (not shown). *E. coli* ecotin has a size of 16 kDa. Thus, it was obvious that the 20 kDa fragment was a result of proteolysis of the full-length fusion protein within the proinsulin region. It was assumed that preparation of the extract from the frozen biomass was unfavorable. Likewise, any time-consuming purification process should be avoided. As will be shown below, the quantity of full-length fusion protein could be significantly improved using fresh biomass obtained after optimized fermentation conditions. A new affinity purification strategy was developed, in order to speed up the purification process.

Improvement of ecotin-proinsulin production

Next, cultivation parameters for the production of ecotin-proinsulin were studied in batch in four 400 ml parallel vessels using *E. coli* BL21(DE3)Gold as a host. The cells were grown in a synthetic medium (M9) containing glucose (25 g/L) as a carbon source. The cells were first grown at 37 °C for 5 hours. To evaluate the influence of the incubation temperature and cultivation time on the yield of fusion protein, the temperature of the vessels was lowered to 35, 30, 25, and 20 °C, respectively, after addition of 1 mM IPTG, and probes were taken at 1 h intervals. The amount of fusion protein present in the periplasmic extract was quantified by ELISA. The cell-specific yield, taking different final optical densities into account, is given in Fig. 2A. A temperature of 25-30 °C and 2-4 hours of incubation were most suitable.

The amount of inducer present in the medium can strongly influence the yield of folded proteins [27]. Thus, the effect of different IPTG concentrations was studied. In parallel vessels, 0.05, 0.1, 0.5, and 1 mM of IPTG, respectively, were added. The temperature was lowered from 37 °C to 25 °C and the cells were further grown for 4 hours. The amount of ecotin-proinsulin determined by ELISA was nearly unchanged, indicating that the inducer concentration was not limiting. Even 0.05 mM IPTG was sufficient (Fig. 2B).

Plasmid stability is another pivotal aspect in the optimization of recombinant protein production. Next, the production of the fusion protein was studied in the presence of different concentrations of ampicillin and its acid-stable analogue, carbenicillin [28]. As shown in Fig. 2C, the effect was rather negligible, and the final yield of fusion protein produced was nearly unchanged. For further experiments ampicillin concentration was kept at 100 mg/L.

High density fed-batch fermentation

The fermentation process was then optimized in a BIOSTAT reactor with a working volume of 8 L under moderate and high cell density conditions. The first fermentations were performed in M9 medium. The feeding solution contained 40% glucose. After an initial cultivation period at 37 °C, fusion protein synthesis was initiated with 1 mM IPTG and the cells were further grown for 3 hours at 25 °C. When the synthesis of ecotin-proinsulin was induced at moderate density (OD_{600} of 66) in M9 medium, 55 mg of fusion protein (18.5 mg proinsulin) were produced per L fermentation broth (Table 1). The fermentation was repeated,

but now IPTG was added at a higher density (OD_{600} of 158). Under these conditions, the yield was reduced to only 27 mg fusion protein per L fermentation broth or to 9.0 mg proinsulin (Table 1). The cell-specific yield, taking the different final densities into account, had decreased more than four-fold (157 µg versus 677 µg fusion protein per L OD1, Table 1).

Further fermentation experiments were done in M9 medium supplemented with peptone (15 g/L). In this medium, the yield was much higher, even when IPTG was added at high density. In the first fermentation using semisynthetic medium, fusion protein synthesis was started at OD_{600} of 161. Here, 153 mg fusion protein (51.5 mg proinsulin) was obtained per L fermentation broth. The experiment was repeated a second time and IPTG was added at OD_{600} of 167. Now, 111 mg/L fusion protein (37.3 mg/L proinsulin) was produced. The cell-specific yield in this medium was also strongly increased (908 and 614 µg fusion protein per L OD1, respectively, as compared to 157 µg fusion protein per L OD1 in M9 medium, table 1). The biomass obtained from the second fermentation was used to purify and characterize the fusion protein, as described below.

The host strain *E. coli* BL21(DE3)Gold contains an intact *relA* gene and expresses ppGpp synthetase I during amino acid starvation [29, 30]. RNA polymerase is a target for ppGpp, and this stringent response is extremely important in plasmid replication as RNA polymerase plays a key role in the initiation of replication of several replicons [31]. It was demonstrated that most of the ColE1-like plasmids are unable to replicate efficiently in a *relA*$^+$ host during starvation for different amino acids [32].

Purification of the ecotin-proinsulin fusion protein

To avoid proteolysis of the fusion protein during freezing and thawing of the biomass, the periplasmic extract was prepared directly just after completion of the second fermentation process. According to need, the small-scale protocol described by Winter *et al* [19] had to be modified to scale up from milliliter to liter scale. To compare the yield under scaling-up conditions, the same biomass was subjected to a small scale and large scale periplasm extraction. The recovery of the fusion protein under large scale conditions was only 63% as compared to the milliliter scale (70 mg fusion protein were obtained per L broth, as compared to 111 mg, see Table 1). Two technical bottlenecks were observed. Sedimentation of cells in the viscous periplasmic extraction buffer containing 25% sucrose was not complete even after

increasing the centrifugation time. Another limiting factor was the low speed of centrifugation (7,000 rpm) for the extraction of proteins from the osmotic shocked cells.

Trypsinogen affinity chromatography

Ecotin is a broad-range serine-protease inhibitor, continuously produced during all stages of cell growth. It was already known that when rat trypsinogen was recombinantly produced and secreted into the periplasm of *E. coli*, it formed a relatively tight and specific complex with *E. coli* ecotin at neutral pH [33]. This complex could be dissolved by acidification. Trypsinogen is a zymogen and has no proteolytic activity. Therefore, trypsinogen might be suited as an affinity material for the efficient enrichment of ecotin fusion proteins present in the periplasmic extract.

To this end, bovine trypsinogen was immobilized on a pre-packed NHS-activated sepharose column. The periplasmic extract obtained by osmotic shock was loaded onto the column at 4 °C and was washed with 100 mM Tris-HCl, pH 8.5. Next, 50 mM HCl was added, and a single symmetrical peak eluted from the column (Fig. 3A). The fractions were immediately neutralized by adding 0.1 volume of 1 M Tris-HCl, pH 8.5, and were analyzed by SDS-PAGE (Fig. 4). A protein of 28 kDa was present, together with two smaller proteins of about 16 and 18 kDa (Lane 4). ELISA data showed that the eluted protein contained native proinsulin.

Next, the fractions containing native proinsulin were pooled and were digested by thrombin for 2 h at room temperature. The full-length fusion protein disappeared, and obviously was converted into GS-H_6R-proinsulin (10.5 kDa) and ecotin-$(GS)_6$-LVPR (17.4 kDa) (Fig 4, lane 5). After thrombin cleavage of the fusion protein, a duplication of the ELISA signal was observed, indicating that in the native homodimeric fusion protein only one of the two proinsulin moieties is accessible to the detection antibody. Ecotin is a homodimeric protein. The fusion protein is also in the dimeric state, with proinsulin moieties sticking at opposite ends. If the fusion protein is analyzed by the sandwich ELISA, proinsulin from one end is bound by the capture antibody immobilized on the plate surface. The second proinsulin moiety at the opposite end of the dimer is free to be recognized by the detection antibody.

Proinsulin liberated after thrombin cleavage still contained the GS-H_6R-moiety at the N-terminal end. Thus, Ni-NTA affinity chromatography was performed to separate ecotin from

proinsulin (Fig. 4B). The thrombin-digested protein solution was loaded onto Ni-NTA resin. Wild-type ecotin (16 kDa) and its recombinant derivative were present in the flow-through (Fig. 4A, lane 6). The remaining protein bound to Ni-NTA was eluted by a linear gradient of 0 – 300 mM imidazole. Two peaks were obtained (Fig. 3B), and were analyzed by ELISA and SDS-PAGE. The vast majority of proinsulin was present in the second peak. SDS-PAGE analysis revealed a homogenous protein of the expected size of 10.5 kDa (Fig. 4B, lane 3). In the first peak only marginal amounts of proinsulin were found. A respective protein band was not visible after SDS-PAGE and staining with Coomassie (Fig. 4B, lane 2). The signal at 280 nm obtained for this peak might be due to small amounts of remaining full-length fusion protein, which should also bind to Ni-NTA but elutes first due to the presence of an internal his-tag which might cause steric hindrance. The molar extinction coefficient of the fusion protein dimer is 10 times higher (ε = 58,760 mol^{-1} cm^{-1}) than that of proinsulin (ε = 6,335 mol^{-1} cm^{-1}), although both give an ELISA signal.

A summary of the affinity purification is given in Table 2. The purification started from a wet biomass of 4.5 g cells with 250 ml of periplasmic extract containing 388 µg of native proinsulin. Almost 48% was eluted from the trypsinogen affinity column, and 43% were present after Ni-NTA chromatography. The final yield was 170 µg proinsulin obtained in the 2^{nd} peak of Ni-NTA. To verify the final yield of purified proinsulin, the amount of protein was also determined with the Micro BCA assay using BSA as a standard (238 µg).

Characterization of the purified fusion protein

Probes obtained during purification were collected and analyzed by mass ESI-MS (Table 3). For the full-length fusion protein a mass of 27,921 Da was determined. The calculated mass for the reduced fusion protein is 9 Da more (27,930 Da). Ecotin contains one disulfide bond, and three disulfide bonds are present in native proinsulin. Thus, the fusion protein was obtained in its native oxidized state. After thrombin digestion, masses of 17,428 Da and 10,511 Da were obtained, corresponding to the oxidized ecotin part [ecotin-$(GS)_6$-LVPR, 17,428 Da] and oxidized proinsulin [GS-H_6R-proinsulin, 10,512 Da]. Wild-type ecotin is naturally localized in the periplasm. It was copurified on the trypsinogen column. The determined mass of 16,098 Da corresponded to its oxidized state.

Effects of the host strain and medium

Proteolysis is one of the major drawbacks for proinsulin production. In the pilot experiment, a 20 kDa proteolytic fragment was found, which might be derived from a cleavage of ecotin-proinsulin at Tyr16-Lys17 or Phe25-Tyr26, a reaction that is catalysed by protease Pi (pitrilysin). *E. coli* BL21(DE3)Gold is devoid of two proteases, one present in the cytoplasm (*lon*$^-$) and one present in the periplasm (*ompT*$^-$), but still contains pitrilysin (*ptr*). *E. coli* SF 120 is devoid of three periplasmic proteases in (*ompT*$^-$, *ptr*$^-$, *degP*$^-$) and was constructed by Meerman and co-workers (1994) [34]. To see whether this protease-deficient host strain might improve the yield of native proinsulin, both strains were grown in shake flasks and the production of ecotin-proinsulin present in the periplasmic extract was analyzed by ELISA after 3 h of induction.

Growth of *E. coli* SF120 was apparently poor in both, LB and M9 medium, with a final OD of 1.3 (LB) and 0.25 (M9 medium), respectively, after 3 hours of protein synthesis at 24 °C. Under the same conditions, *E. coli* BL21(DE3)Gold reached OD values of 4 (LB), and 0.8 (M9 medium). Taking the different final optical densities into account, the amount of fusion protein produced in SF120 was higher (85 and 79 µg/L OD1) than in BL21(DE3)Gold (21 and 23 µg/L OD1, respectively; Fig. 5A). However, due to the poor growth of SF210, the volumetric yield of ecotin-proinsulin in this strain was only marginally increased (Fig. 5B). Moreover, after growth of SF210 under high density fermentation conditions, no ecotin-proinsulin fusion protein was detectable.

Concluding remarks

In this study, human proinsulin was produced in the bacterial periplasm as a fusion to *E. coli* ecotin. The protein was produced in its native state, as was evident from the ELISA, in which only native insulin or native proinsulin with correct disulfide patterns are recognized. The native oxidized state was further confirmed by mass spectrometry of the purified protein.

After high cell density fermentation (OD_{600} of 170) in *E. coli* BL21(DE3) 153 mg fusion protein (ecotin-proinsulin) per L fermentation broth were produced, corresponding to 51.1 mg proinsulin. In shake flask experiments using the same strain, the amount of fusion protein produced was relatively low (about 0.1 mg fusion protein per L broth, see Fig. 5B). The increase in yield under high-density fermentation conditions is not only due to the higher cell density, since the cell-specific yield also increased under these conditions (0.02 mg fusion protein/L OD1 under shake flask conditions, and 0.9 mg fusion protein/L OD1 during high-density fermentation). As analyzed in batch fermentations, the influence of the inducer and antibiotic concentration on the yield of fusion protein were rather negligible. The incubation temperature was more important, with 25 °C being the best, which was also published before [19]. The medium composition, however, played a crucial role, at least under fermentation conditions. Addition of peptone (15 g/L) under high-density conditions resulted in a nearly five-fold increase in the yield of fusion protein (153 and 111 mg/L as compared to 27 mg/L).

In another fusion system for the production of native proinsulin in the periplasm, single and double Z domains of staphylococcal protein A were fused at the N-terminus of proinsulin. Experiments were performed in shake flasks with *E. coli* JM109 and AF1000 as host strains using LB and M9 medium. The best yield was 7.2 mg/g, using *E. coli* JM109 in shake flasks. [18]. This would on average correspond to 5-10 mg ZZ-proinsulin per L broth, assuming an optical density of 5 and 1 g dry weight per L medium. It was suggested that acidic region of the ZZ domain might interact with the basic residues of proinsulin via electrostatic interactions thus protecting the fusion protein from proteolysis [18]. In an earlier study, using the same fusion system (ZZ-proinsulin), the length of connecting peptide of proinsulin was varied [14]. It was found that when most of the connecting peptide (24 out of 35 residues) was removed, the yield of native proinsulin analogue was increased more than 25 fold, reaching up to 58 mg/L in 2 x YT medium. The results indicated that the higher yield of proinsulin in the periplasm was due to an improvement in export efficiency [14].

When proinsulin was fused to the C-terminus of full-length DsbA, 0.11 mg/L native proinsulin was obtained in shake flask experiments using *E. coli* BL21 as a host [19]. Winter *et al.* compared the yield of DsbA-proinsulin in other host strains (SF131, JM109, C600 and RB791). The volumetric yield of proinsulin was nearly 2-fold (0.24 mg/L) in C600 and 3-fold higher (0.29 mg/L) in RB791. The effect of medium additives during production of DsbA-proinsulin in shake flask was also studied [19]. The cell-specific yield of DsbA-proinsulin was 3.4 fold increased when the medium was supplemented with 0.4 M L-Arginine due to suppression of protein aggregation [35, 36]. However, due to a negative effect on cell growth, the volumetric yield did not increase. Addition of 1 % (v/v) ethanol, 1 % yeast extract in minimal medium lead to an increase in the specific yield of DsbA-proinsulin by a factor of 2.3 and 1.3, respectively. These might be potential improvement factors to increase the yield of native ecotin-proinsulin in the future.

A major disadvantage of the periplasm as a folding compartment for proteins with multiple disulfide bonds is the presence of the strong oxidant DsbA. It has been shown that DsbA introduces disulfide bonds into translocating polypeptides as soon as two cysteines have emerged into the periplasm [37, 38]. Although disulfide isomerases exist in the periplasm, these are often not sufficient to correct wrongly paired disulfides of proteins, especially if they contain multiple disulfide bonds of nonlinear connectivities [39]. The consequence is inclusion body formation in the periplasm [40]. Schäffner *et al.* [23] designed a strategy to suppress inclusion body formation in the periplasm by adding disulfide reshuffling reagents and substances known to stabilize folding intermediates to the medium. Also, the effect of cosecretion of the ATP-independent molecular chaperone DnaJ was analyzed [23]. DnaJ has been shown to suppress aggregation of non-native proteins *in vitro* [41]. Under optimal conditions, cosecretion of DnaJ into the periplasm increased the yield of native proinsulin by a factor of 37 [23]. In another study, DsbC was overexpressed, and the yield of native recombinant proteins in the periplasm was considerably increased [42].

We observed severe proteolysis of the fusion protein in the proinsulin region when the periplasmic extract was prepared from a frozen biomass. Freezing and thawing can promote cell lysis, and the cytoplasmic content is released during periplasm extraction. The half-life of proinsulin in the cytoplasm is only 2 min as compared to 20 min in the periplasm [20]. Proinsulin is relatively small in size; it contains many basic residues and is thus an inherently

good substrate for proteases. Insulin is a well-know substrate for at least two metalloproteases (Ci and Pi) of *E. coli* [43]. Several strategies were suggested for the protection of recombinant proteins from proteolysis [44-46]. Some promising solutions for minimizing proteolysis included sequence modifications, upstream fusion [47], use of protease deficient strains [34]. Likewise, the temperature, inducer and pH-conditions were analyzed [48, 49], and a supplementation with amino acids was tested [50]. In addition, extreme growth limitation in high-density fed-batch cultivation was avoided [51, 52]. Another straightforward approach to reduce the proteolysis should be the use of protease deficient strains. However, the growth of these strains is often very poor, and the protease activity may be recovered via by-pass mechanism [34, 53]. Poor growth of strain SF120 (*ompT*$^-$, *ptr*$^-$, *degP*$^-$) was also observed in this study.

The yield of the ecotin-proinsulin fusion protein could strongly be increased by optimizing the fermentation process. The ecotin-proinsulin fusion protein was designed to efficiently translocate proinsulin into the periplasm, to improve folding and to easily purify it (Fig. 1). The new protocol using a trypsinogen affinity column allows a purification of the ecotin-proinsulin fusion protein within a couple of hours at low temperature without dialysis. This substantially improves the quality of the fusion protein. It is thus suitable for proteins that are likely to be degraded very fast. Ecotin is a broad-range serine protease inhibitor that tightly binds these proteases. Lengyel *et al.* [33] found that even a trypsin precursor (trypsinogen) has a high affinity to ecotin. Any protein linked to C-terminus of ecotin would always be perpendicular to the macromolecular trypsinogen-ecotin complex. The trypsinogen affinity purification was quite efficient, as only the respective fusion protein was bound and eluted. Consequently, proteases present in the periplasmic extract were removed at an early purification stage, and 48% of the fusion protein present in the periplasmic extract could be recovered after trypsinogen affinity chromatography. Trypsinogen immobilized on the column was fairly stable and more than 70% of the binding capacity was retained after 16 rounds of acid elution. Trypsinogen as a capture molecule for the purification of ecotin-proinsulin should not cause a degradation of the fusion protein. However, if the thrombin-cleaved fusion protein was passed again through the trypsinogen column to remove the ecotin part, proteolytically processed proinsulin was present in the flow-through (data not shown). Traces of trypsin present in the commercial trypsinogen preparation might be an explanation. Trypsin removes a major part of C-peptide from proinsulin and converts it to an insulin like molecule [11]. In future, this affinity step should be further improved by using an

enzymatically inactive trypsin mutant. One of the undesirable effects of periplasmic extraction is the large volume, obtained by the osmotic shock procedure. Using the immobilized trypsinogen or an inactive trypsin variant, binding could be performed in wide range of conditions keeping neutral/slight alkaline pH.

In an earlier study [8], several different signal sequences of prokaryotic origin (*pelB*, *ompT* and *dsbA*) were fused to the N-terminus of human pepsinogen to direct it into the periplasm, but no native protein was obtained. A fusion of pepsinogen to the C-terminus of *E. coli* ecotin proved to be a solution [8]. The use of ecotin as an N-terminal fusion tag has several advantages. Ecotin has no role in the host metabolism. Due to its relatively small size, the metabolic burden is low. After secretion into the periplasm, it does not interact with other *E. coli* proteins [54]. Thus, no adverse effect upon over-expression is expected. Fusion proteins can be quantified using the ecotin moiety by a trypsin inhibition assay [55]. Ecotin is a dimeric protein. Gel permeation data showed that also a fusion protein is present as dimer [8]. Due to domain swapping, the C-termini of ecotin stick to opposite directions, and the folding of fusion partners takes place at each ends without steric hindrance. Furthermore, the presence of a disulfide bond in ecotin implies an oxidative folding pathway. Ecotin has a very stable fold that might help to stabilize a C-terminally fused protein. This also helps to purify the desired protein. Recently it was found that ecotin is native, even when expressed in the cytoplasm of *E. coli* [55]. Thus, it might also be useful as folding enhancer for cytosolic expression.

Acknowledgments

The present work was supported by a grant from the Stiftung Industrieforschung, Germany, and by grants from the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) of Germany, and from the Land Sachsen-Anhalt. The authors thank A. Schierhorn, Halle, for mass spectrometry, and P. Rücknagel, Halle, for N-terminal sequencing.

References

[1] F. Baneyx, Recombinant protein expression in *Escherichia coli*, Curr. Opin. Biotechnol. 10 (1999) 411-421.

[2] O. Pines, M. Inouye, Expression and secretion of proteins in *E. coli*, Mol. Biotechnol. 12 (1999) 25-34.

[3] O. Francetic, D. Belin, C. Badaut, A. P. Pugsley, Expression of the endogenous type II secretion pathway in *Escherichia coli* leads to chitinase secretion, EMBO J. 19 (2000) 6697-6703.

[4] A. Economou, Following the leader: bacterial protein export through the Sec pathway, Trends Microbiol. 7 (1999) 315-320.

[5] A. P. Pugsley, O. Francetic, A. J. Driessen, V. de Lorenzo, Getting out: protein traffic in prokaryotes, Mol. Microbiol. 52 (2004) 3-11.

[6] A. V. Kajava, S. N. Zolov, A. E. Kalinin, M. A. Nesmeyanova, The net charge of the first 18 residues of the mature sequence affects protein translocation across the cytoplasmic membrane of gram-negative bacteria, J. Bacteriol. 182 (2000) 2163-2169.

[7] O. V. Khokhlova, M. A. Nesmeyanova, Interaction of SecB and SecA with the N-terminal region of mature alkaline phosphatase on its secretion in *Escherichia coli*, Mol. Biol. 37 (2003) 605-610.

[8] A. Malik, R. Rudolph, B. Sohling, A novel fusion protein system for the production of native human pepsinogen in the bacterial periplasm, Protein Expr. Purif. 47 (2006) 662-671.

[9] C. H. Chung, H. E. Ives, S. Almeda, A. L. Goldberg, Purification from *Escherichia coli* of a periplasmic protein that is a potent inhibitor of pancreatic proteases, J. Biol. Chem. 258 (1983) 11032-11038.

[10] K. Docherty, R. J. Carroll, D. F. Steiner, Conversion of proinsulin to insulin: involvement of a 31,500 molecular weight thiol protease, Proc. Natl. Acad. Sci. USA 79 (1982) 4613-4617.

[11] W. Kemmler, J. D. Peterson, D. F. Steiner, Studies on the conversion of proinsulin to insulin. I. Conversion in vitro with trypsin and carboxypeptidase B, J. Biol. Chem. 246 (1971) 6786-6791.

[12] I. S. Johnson, Human insulin from recombinant DNA technology, Science 219 (1983) 632-637.

[13] G. Walsh, Therapeutic insulins and their large-scale manufacture, Appl. Microbiol. Biotech. 67 (2005) 151-159.

[14] Y. Kang, J. W. Yoon, Effect of modification of connecting peptide of proinsulin on its export, J. Biotechnol 36 (1994) 45-54.

[15] J. G. Tang, M. H. Hu, Production of human proinsulin in *Escherichia coli* in a non-fusion form, Biotechnol. Lett. 15 (1993) 661-666.

[16] Y. Kang, J. W. Yoon, Development of a high-expression vector (Pyk10-9) of human proinsulin gene, Biotechnol. Lett. 13 (1991) 755-760.

[17] W. L. Sung, F. L. Yao, D. M. Zahab, S. A. Narang, Short synthetic oligodeoxyribonucleotide leader sequences enhance accumulation of human proinsulin synthesized in *Escherichia coli*, Proc. Natl. Acad. Sci. USA 83 (1986) 561-565.

[18] F. J. Mergulhao, M. A. Taipa, J. M. Cabral, G. A. Monteiro, Evaluation of bottlenecks in proinsulin secretion by *Escherichia coli*, J. Biotechnol. 109 (2004) 31-43.

[19] J. Winter, P. Neubauer, R. Glockshuber, R. Rudolph, Increased production of human proinsulin in the periplasmic space of *Escherichia coli* by fusion to DsbA, J. Biotechnol. 84 (2000) 175-185.

[20] K. Talmadge, W. Gilbert, Cellular location affects protein stability in *Escherichia coli*, Proc. Natl. Acad. Sci. USA 79 (1982) 1830-1833.

[21] K. Talmadge, J. Brosius, W. Gilbert, An internal signal sequence directs secretion and processing of proinsulin in bacteria, Nature 294 (1981) 176-178.

[22] S. J. Chan, J. Weiss, M. Konrad, T. White, C. Bahl, S. D. Yu, D. Marks, D. F. Steiner, Biosynthesis and periplasmic segregation of human proinsulin in *Escherichia coli*, Proc. Natl. Acad. Sci. USA 78 (1981) 5401-5405.

[23] J. Schaffner, J. Winter, R. Rudolph, E. Schwarz, Cosecretion of chaperones and low-molecular-size medium additives increases the yield of recombinant disulfide-bridged proteins, Appl. Environ. Microbiol. 67 (2001) 3994-4000.

[24] A. S. Rathore, R. E. Bilbrey, D. E. Steinmeyer, Optimization of an osmotic shock procedure for isolation of a protein product expressed in *E. coli*, Biotechnol. Prog. 19 (2003) 1541-1546.

[25] M. Jenzsch, S. Gnoth, M. Kleinschmidt, R. Simutis, A. Lubbert, Improving the batch-to-batch reproducibility in microbial cultures during recombinant protein production by guiding the process along a predefined total biomass profile, Bioprocess Biosyst. Eng. 29 (2006) 315-321.

[26] C. N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, How to Measure and Predict the Molar Absorption-Coefficient of a Protein, Protein Sci. 4 (1995) 2411-2423.

[27] P. Turner, O. Holst, E. N. Karlsson, 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 (2005) 54-60.

[28] H. P. Sorensen, K. K. Mortensen, Advanced genetic strategies for recombinant protein expression in *Escherichia coli*, J. Biotechnol 115 (2005) 113-128.

[29] M. Jishage, K. Kvint, V. Shingler, T. Nystrom, Regulation of sigma factor competition by the alarmone ppGpp, Genes Dev. 16 (2002) 1260-1270.

[30] D. Joseleau-Petit, D. Vinella, R. D'Ari, Metabolic alarms and cell division in *Escherichia coli*, J. Bacteriol. 181 (1999) 9-14.

[31] M. M. Barker, T. Gaal, C. A. Josaitis, R. L. Gourse, Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro, J. Mol. Biol. 305 (2001) 673-688.

[32] B. Wrobel, G. Wegrzyn, Replication regulation of ColE1-like plasmids in amino acid-starved *Escherichia coli*, Plasmid 39 (1998) 48-62.

[33] Z. Lengyel, G. Pal, M. Sahin-Toth, Affinity purification of recombinant trypsinogen using immobilized ecotin, Protein Expr. Purif. 12 (1998) 291-294.

[34] H. J. Meerman, G. Georgiou, Construction and characterization of a set of *E. coli* strains deficient in all known loci affecting the proteolytic stability of secreted recombinant proteins, Bio/Technology 12 (1994) 1107-1110.

[35] K. Tsumoto, D. Ejima, Y. Kita, T. Arakawa, Review: Why is arginine effective in suppressing aggregation, Prot. Pept. Lett. 12 (2005) 613-619.

[36] E. D. Clark, E. Schwarz, R. Rudolph, Inhibition of aggregation side reactions during in vitro protein folding, Methods Enzymol 309 (1999) 217-236.

[37] J. C. Joly, J. R. Swartz, In vitro and in vivo redox states of the *Escherichia coli* periplasmic oxidoreductases DsbA and DsbC, Biochemistry 36 (1997) 10067-10072.

[38] M. Sone, Y. Akiyama, K. Ito, Differential in vivo roles played by DsbA and DsbC in the formation of protein disulfide bonds, J. Biol. Chem. 272 (1997) 10349-10352.

[39] D. Missiakas, S. Raina, Protein folding in the bacterial periplasm, J. Bacteriol. 179 (1997) 2465-2471.

[40] G. Georgiou, J. N. Telford, M. L. Shuler, D. B. Wilson, Localization of inclusion bodies in *Escherichia coli* overproducing beta-lactamase or alkaline phosphatase, Appl. Environ. Microbiol. 52 (1986) 1157-1161.

[41] H. Schroder, T. Langer, F. U. Hartl, B. Bukau, DnaK, DnaJ and GrpE form a cellular chaperone machinery capable of repairing heat-induced protein damage, EMBO J. 12 (1993) 4137-4144.

[42] J. Qiu, J. R. Swartz, G. Georgiou, Expression of active human tissue-type plasminogen activator in *Escherichia coli*, Appl. Environ. Microbiol. 64 (1998) 4891-4896.

[43] S. Gottesman, Proteases and their targets in *Escherichia coli*, Annu. Rev. Genet. 30 (1996) 465-506.

[44] S. C. Makrides, Strategies for achieving high-level expression of genes in *Escherichia coli*, Microbiol. Rev. 60 (1996) 512-538.

[45] G. Hannig, S. C. Makrides, Strategies for optimizing heterologous protein expression in *Escherichia coli*, Trends Biotechnol. 16 (1998) 54-60.

[46] C. Herman, R. D'Ari, Proteolysis and chaperones: the destruction/reconstruction dilemma, Curr. Opin. Microbiol. 1 (1998) 204-209.

[47] M. Murby, M. Uhlen, S. Stahl, Upstream strategies to minimize proteolytic degradation upon recombinant production in *Escherichia coli*, Protein Expr. Purif. 7 (1996) 129-136.

[48] Q. H. Yang, C. L. Wu, K. Lin, L. Li, Low concentration of inducer favors production of active form of 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase in *Escherichia coli*, Protein Expr. Purif. 10 (1997) 320-324.

[49] F. Baneyx, A. Ayling, T. Palumbo, D. Thomas, G. Georgiou, Optimization of growth conditions for the production of proteolytically-sensitive proteins in the periplasmic space of *Escherichia coli*, Appl. Microbiol. Biotechnol. 36 (1991) 14-20.

[50] D. M. Ramirez, W. E. Bentley, Fed-batch feeding and induction policies that improve foreign protein synthesis and stability by avoiding stress responses, Biotechnol. Bioeng. 47 (1995) 596-608.

[51] A. C. St John, A. L. Goldberg, Effects of starvation for potassium and other inorganic ions on protein degradation and ribonucleic acid synthesis in *Escherichia coli*, J. Bacteriol. 143 (1980) 1223-1233.

[52] K. L. Larrabee, J. O. Phillips, G. J. Williams, A. R. Larrabee, The relative rates of protein synthesis and degradation in a growing culture of *Escherichia coli*, J. Biol. Chem. 255 (1980) 4125-4130.

[53] F. Baneyx, G. Georgiou, Construction and characterization of *Escherichia coli* strains deficient in multiple secreted proteases: protease III degrades high-molecular-weight substrates in vivo, J. Bacteriol. 173 (1991) 2696-2703.

[54] C. T. Eggers, L. A. Murray, V. A. Delmar, A. G. Day, C. S. Craik, The periplasmic serine protease inhibitor ecotin protects bacteria against neutrophil elastase, Biochemical J. 379 (2004) 107-118.

[55] S. H. Kang, J. J. Park, S. S. Chung, O. S. Bang, C. H. Chung, Strategies for assaying deubiquitinating enzymes, Methods Enzymol. 398 (2005) 500-508.

Table 1: Cultivation conditions and yield of ecotin-proinsulin under high-density fermentation conditions

Medium	Cell density at time of induction[a]	Cell density at time of harvesting[a]	Ecotin-Proinsulin (mg L^{-1})	Ecotin-Proinsulin (µg $L^{-1}OD1^{-1}$)
M9	66	82	55	677
M9	158	172	27	157
M9+Peptone	161	170	153	908
M9+Peptone	167	182	111	614

[a]The cell density was determined at 600 nm.

Table 2: Purification of proinsulin after periplasmic expression of the ecotin-proinsulin fusion protein

Step	Vol. (ml)	Proinsulin (µg/ml)[b]	Total Proinsulin (µg)[b]	Yield (%)
Periplasmic extract[a]	250	1.55	388	100
Trypsinogen affinity chromatography, including thrombin cleavage	5	37.4	187	48
Flow-through Ni-NTA	15	0.008	0.13	0.03
1st Peak Ni-NTA	8	2.48	19.9	5
2nd Peak Ni-NTA	16	10.6[b], 14.9[c]	170[b], 238[c]	43

[a], The periplasmic extract was prepared from 4.5 g wet biomass of BL21(DE3) pEG-PI.

[b], Proinsulin was determined by ELISA.

[c], Purified proinsulin was determined by the Micro BCA Protein determination kit.

Table 3: ESI-MS analysis of probes after affinity purification of the ecotin-proinsulin fusion protein

Probe	Calculated reduced mass (Da)	ESI-MS (Da)	Difference (Da)
Ecotin-proinsulin	27929.7	27921.0	8.7
Ecotin[a]	17429.9	17428.0	1.9
GS-H_6-R-proinsulin[a]	10517.8	10511.0	6.8
Endogenous ecotin	16099.5	16097.0	2.5

[a], These probes were obtained after digestion with thrombin

Figure Legends

Fig.1. Schematic presentation of the ecotin-proinsulin fusion protein cloned in pTrc99a
Full-length ecotin (1-162) including its signal sequence (SS, 1-20) was fused to the N-terminus of human proinsulin (188-273). A Gly-Ser linker (163-174) of 6 repeats $(GS)_6$, a thrombin cleavage site (LVPRGS; 175-180), a Hexa-histidine tag (181-186) and an arginine (187) residue (R) for trypsin cleavage was introduced in between ecotin and proinsulin. Respective protease cleavage sites are indicated by arrows.

Fig.2. Parameter analysis in batch fermentations for the production of ecotin-proinsulin
(A) Influence of temperature and harvesting time on the yield of fusion protein. After addition of 1 mM IPTG, the temperature was lowered from 37 °C to 35 °C (▲), 30 °C (■), 25 °C (♦) and 20 °C (●). (B) Influence of the IPTG concentration. Protein synthesis was started by addition of 50 µM (▲), 100 µM (■), 500 µM (♦) and 1000 µM IPTG (●). The temperature was lowered form 37 °C to 25 °C. (C) Analysis of the effect of different antibiotics. The fermentation was done in the presence of 100 mg/L carbenicillin (▲), 100 mg/L carbenicillin and ampicillin each (■), 500 mg/L ampicillin (♦), and 100 mg/L ampicillin (●), respectively. The temperature was lowered from 37 °C to 25 °C after addition of 200 µM IPTG.

Fig.3. Affinity purification of ecotin-proinsulin by immobilized trypsinogen
(A) Elution profile of ecotin-proinsulin from the trypsinogen column. The crude periplasmic extract in 10 mM Tris-HCl, 0.5 mM MgCl2, 1 mM EDTA was loaded onto the column. The column was washed with 100 mM Tris-HCl, pH 8.5. The fusion protein was eluted by acidification. The dashed line (---) shows the addition of 50 mM HCl, the solid line represents the protein elution. Fractions were immediately neutralized with 100 mM Tris pH 8.5, and were pooled for subsequent thrombin digestion. (B) After thrombin digestion, the protein solution was subjected to Ni-NTA chromatography. The protein was loaded onto the column in 300 mM NaCl, 5 mM imidazole. The column was washed with 50 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, pH 8.0. His-tagged proinsulin was eluted by a linear gradient of 300 mM imidazole (dashed line, ---). The solid line represents the protein. Peaks 1 and 2 were pooled separately for qualitative and quantitative analysis.

Fig.4. Purification of ecotin-proinsulin via immobilized trypsinogen and Ni-NTA chromatography
The proteins were separated on NuPAGE (4-12%) Bis-Tris/MES gel. (A) Lane 1, marker; lane 2, periplasmic extract; lane 3, Flow-through from the trypsinogen column; lane 4, pooled eluted peak from the trypsinogen column; lane 5, thrombin digestion of the pooled peak; lane 6, flow-through after loading of the thrombin digested proteins onto the Ni-NTA column. (B) Elution of His-tagged proinsulin from Ni-NTA resin. Lane 1, marker, lane 2, 1st peak and lane 3, 2nd peak.

Fig.5. Yield of ecotin-proinsulin in shake flask experiments in *E. coli* BL21(DE3)Gold and in the protease-deficient strain *E. coli* SF120
(A) Specific yield of ecotin-proinsulin in *E. coli* BL21(DE3)Gold and *E. coli* SF120 after cultivation in rich (LB) and minimal (M9) medium. (B) Volumetric yield of ecotin-proinsulin in *E. coli* BL21(DE3)Gold and *E. coli* SF120 after cultivation in rich (LB) and minimal (M9) medium. “Gold” represents *E. coli* BL 21(DE3) Gold; “SF120” stands for *E. coli* SF120; LB, represents LB (Luria-Bertani) medium and (M9) synthetic minimal medium.

Fig. 1

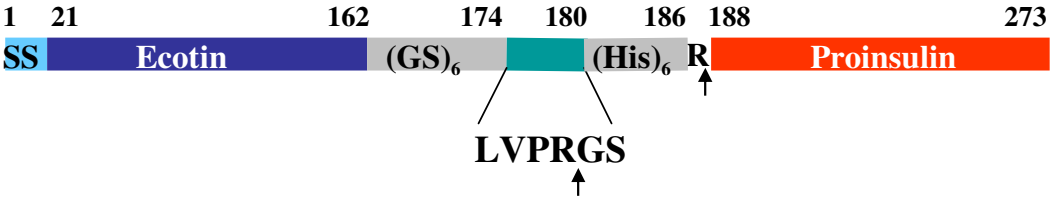

1
21
162
174
180
186
188
273
SS
Ecotin
(GS)6
(His)6
R
Proinsulin
LVPRGS

Fig. 2

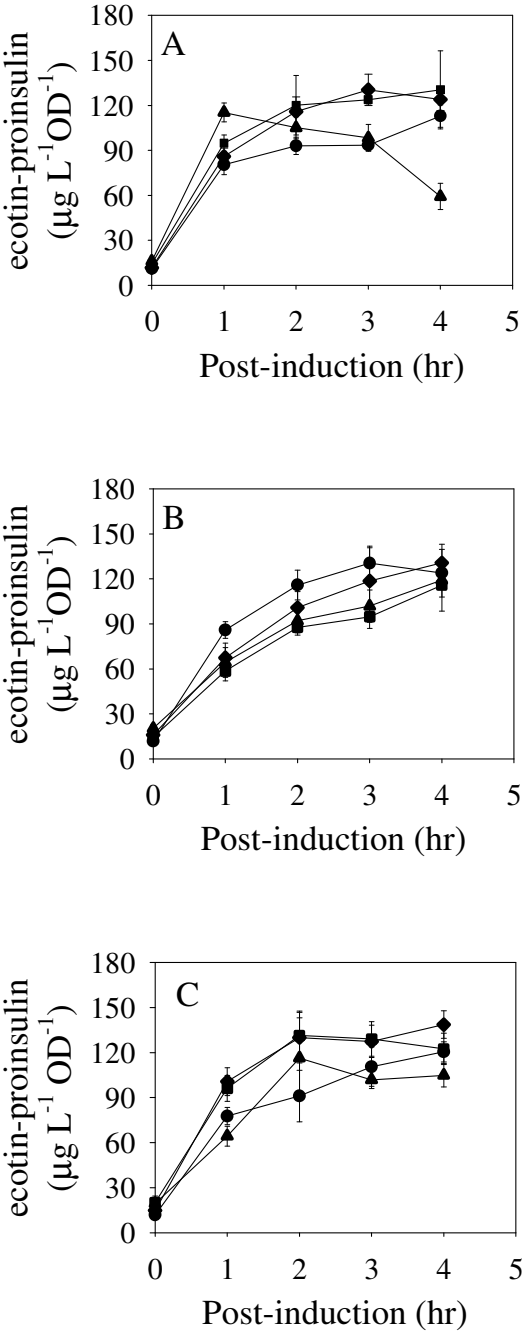

A
B
C
ecotin-proinsulin (μg L-1OD-1)
Post-induction (hr)
0
30
60
90
120
150
180
0
1
2
3
4
5

Fig. 3

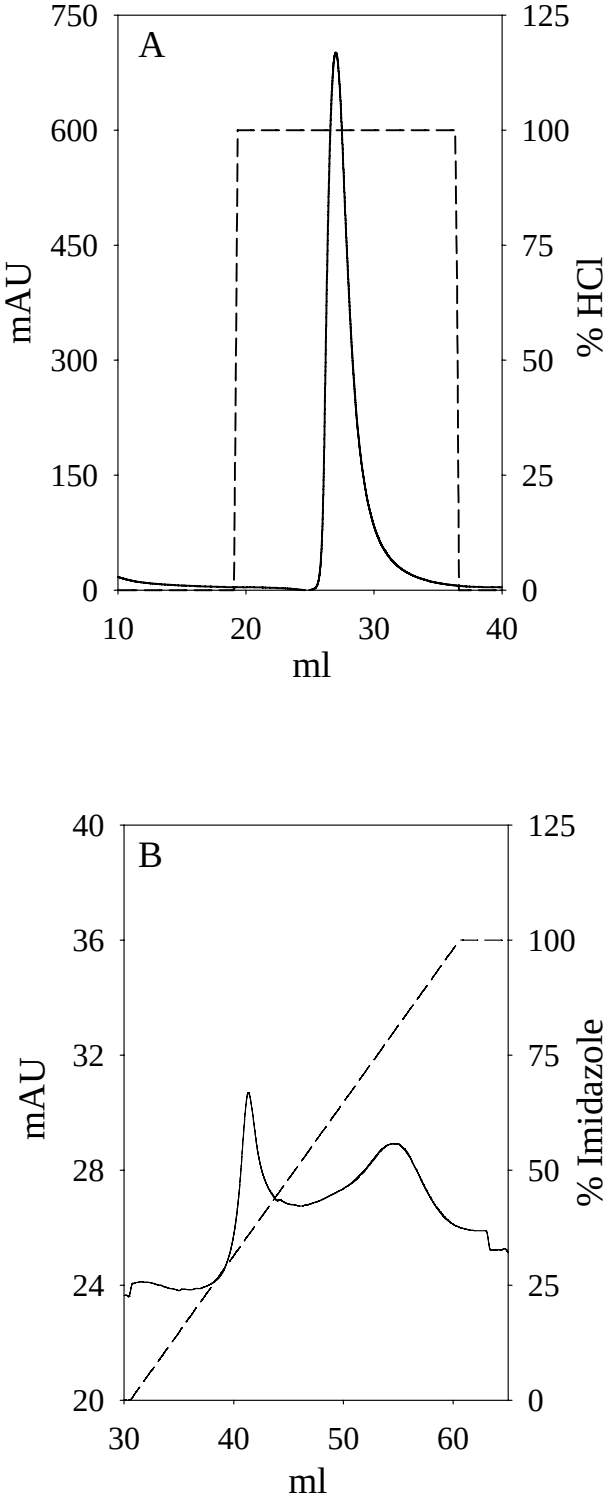
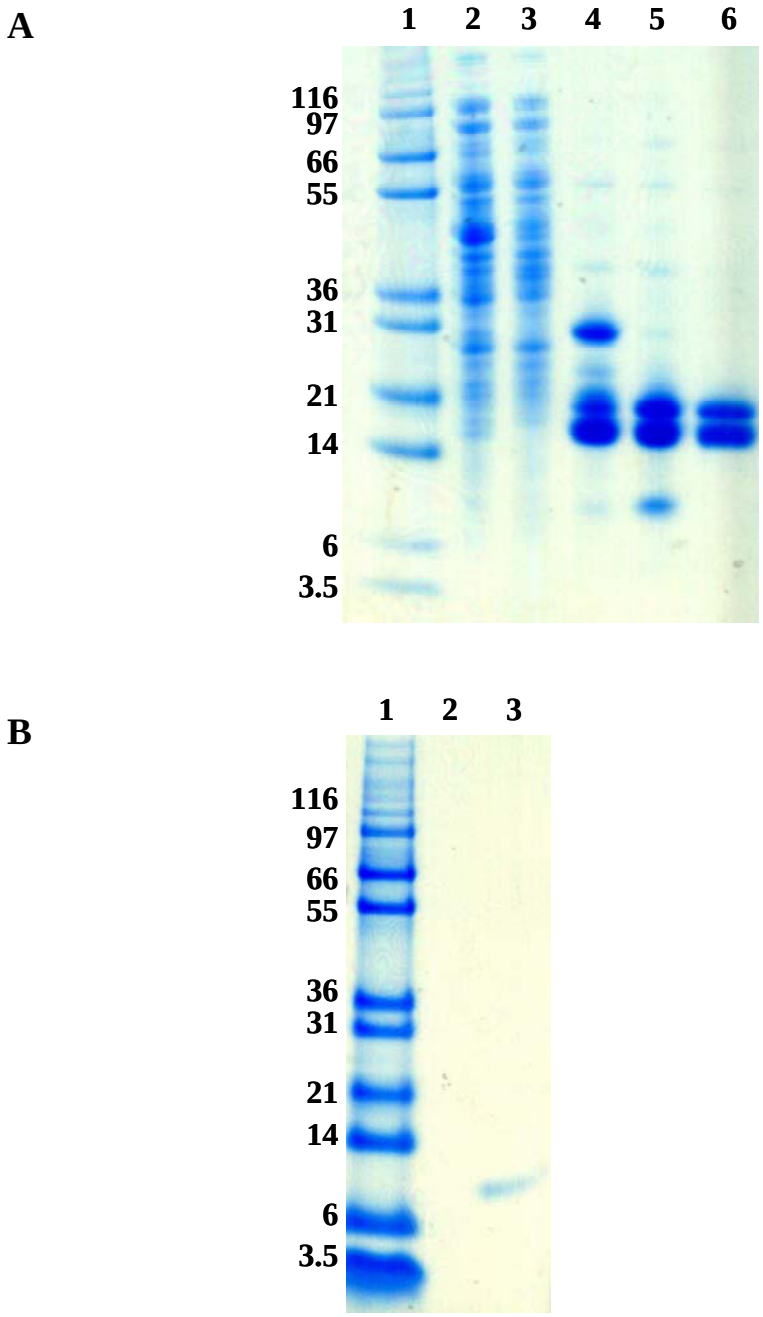
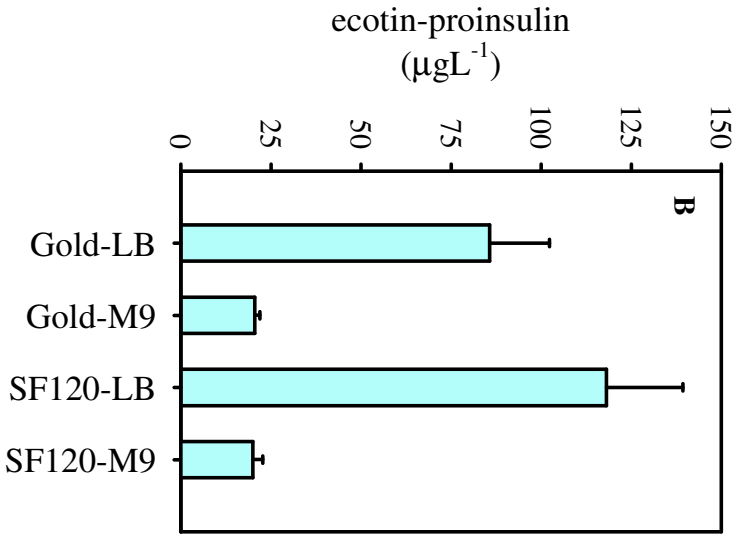

A
750
600
450
300
150
0
mAU
125
100
75
50
25
0
% HCl
10
20
30
40
ml
B
40
36
32
28
24
20
mAU
125
100
75
50
25
0
% Imidazole
30
40
50
60
ml

Fig. 4


A
1 2 3 4 5 6
116
97
66
55
36
31
21
14
6
3.5
B
1 2 3
116
97
66
55
36
31
21
14
6
3.5


ecotin-proinsulin
(µgL-1)
0
25
50
75
100
125
150
B
Gold-LB
Gold-M9
SF120-LB
SF120-M9

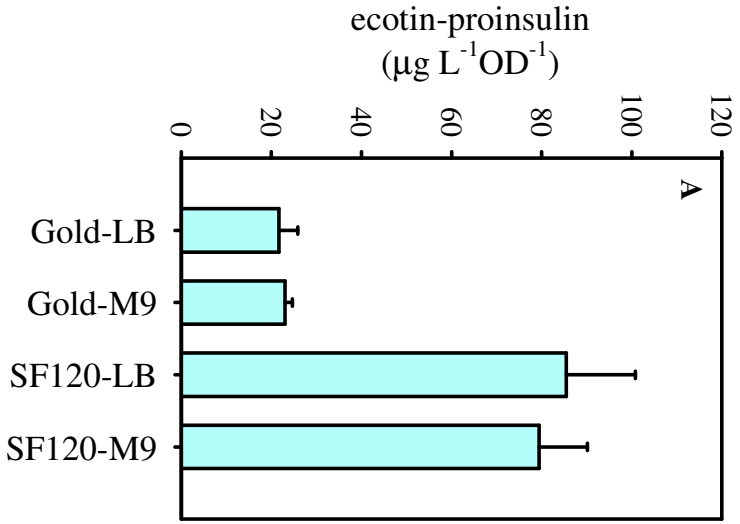

ecotin-proinsulin
(µg L-1OD-1)
0
20
40
60
80
100
120
A
Gold-LB
Gold-M9
SF120-LB
SF120-M9

Fig. 5

Periplasmatische Produktion therapeutisch relevanter Proteine durch Fusion an *Escherichia coli* Ecotin

Anmelder:

Martin-Luther-Universität Halle Wittenberg, 06108 Halle, DE

Erfinder:

Prof. Dr. Rainer Rudolph, Halle, DE
Dr. Brigitte Söhling, Halle, DE
Ajamaluddin Malik, Halle, DE

Anmeldetag: 22.10.2005

Az: 10 2005 050 732.8-41

Veröffentlichung vorraussichtlich am 26.04.2007

Zusammenfassung

Die vorliegende Erfindung betrifft ein Verfahren zur Herstellung natürlich gefalteter rekombinanter Proteine zum Einsatz in der medizinischen Diagnostik und Therapie in einem prokaryontischen Wirtsorganismus. Der Wirtsorganismus enthält eine rekombinante DNA, die für ein Fusionsprotein codiert. Der N-terminale Teil des Fusionsproteins codiert für das Ecotin-Signalpeptid aus Escherichia coli und Ecotin. Der C-terminale Teil des Fusionsproteins enthält die Sequenz des gewünschten rekombinanten Proteins. Die N- und C-terminalen Proteindomänen können dabei durch eine kurze synthetische Peptidsequenz miteinander verbunden sein. Während der Produktion in bakteriellen Wirtszellen wird das Fusionsprotein ins Periplasma sezerniert und es bildet sich die natürliche gefaltete Struktur aus.

Periplasmatische Produktion therapeutisch relevanter Proteine durch Fusion an *Escherichia coli* Ecotin

Die Notwendigkeit einer industriellen Produktion von rekombinanten Proteinen für die Diagnostik und Therapie ist in den letzten beiden Jahrzehnten stetig gewachsen. Dafür werden Techniken benötigt, mit denen ein Protein in der natürlich gefalteten, biologisch aktiven Form gewonnen werden kann. Viele der in der Therapie benötigten Proteine sind sezernierte Proteine mit mehreren Disulfidbrücken. Für diese Proteine ist die Ausbildung der natürlichen Struktur bei der Herstellung oft problematisch. Unter den reduzierenden Bedingungen in bakteriellen Wirtsorganismen fallen diese Proteine in der Regel als Einschlusskörper (inclusion bodies) aus. Es müssen daher geeignete Rückfaltungsverfahren etabliert werden, die die *in vitro* Renaturierung erlauben. Andernfalls kann alternativ eine Herstellung im bakteriellen Periplasma, ein von der inneren und äußeren Membran begrenzter Raum in Gram-negativen Bakterien, erfolgen.

Im oxidierenden Milieu dieses Kompartiments werden die Disulfidbrücken geknüpft, so dass die sekretierten Proteine eine natürlich gefaltete, native Struktur ausbilden. Die Freisetzung aus dem Periplasma erfolgt durch osmotischen Schock mit anschließender Aufreinigung. Eine Produktion der natürlich gefalteten Proteine in eukaryontischen Wirtsorganismen ist vergleichsweise aufwendig und kostenintensiv.

Für die Sekretion ins Periplasma ist ein Translokationssignal erforderlich, eine kurze bakterielle N-terminale Signalsequenz, die nach Translokation durch wirtseigene periplasmatische Signalpeptidasen abgespalten wird (Danese und Silhavy, 1998). Jedoch werden viele Proteine mittels der kurzen bakteriellen Signalpeptide nicht oder nur unzureichend sekretiert. So ist bekannt, dass stark basische N-termini mit mehreren Arginin-Resten die Translokation verhindern (Summers *et al.*, 1989).

Es ist möglich, dieses Sekretionsproblem durch die Verwendung von Fusionsproteinen zu umgehen. Hierbei wird das rekombinante Protein mit einem bakteriellen periplasmatischen Protein oder einer N-terminalen Domäne desselben fusioniert. Proteinfusionen werden in der rekombinanten Proteinproduktion sehr häufig angewandt. Es stehen jedoch wenig Fusionsproteine für die Produktion im Periplasma zur Verfügung.

Bekannt als Fusionsparter für verbesserte Sekretion und Löslichkeit rekombinanter Proteine ist das Maltose-Bindeprotein aus *E. coli* MalE (di Guana *et al.*, 1988). Das MalE Proprotein wird im Cytoplasma vom Sekretionsfaktor SecB in ungefaltetem Zustand gehalten (Liu *et al.*, 1989). Mit MalE fusionierte Zielproteine zeigen daher verbesserte Sekretionseigenschaften. MalE selbst hat jedoch bereits eine Masse von 41 kDa, die bei der Fusion mit Proteinen großer Masse zu Schwierigkeiten führen kann.

Ein weiterer bekannter Fusionspartner ist das *Staphylococcus* Protein A (31 kDa), welches alternativ als Fusionspartner eingesetzt werden kann. Bei Expression in Gram-positiven Wirtsorganismen wird das Fusionsprotein in das umgebende Medium sezerniert, für die Expression in *Escherichia coli* wurden cytoplasmatische und periplasmatische Varianten verwendet (Nilsson *et al.*, 1985). Eine verkürzte Variante (14 kDa), codiert nur für das Signalpeptid sowie zwei synthetische IgG-bindende Z-Domänen (Moks *et al.*, 1987). Das entsprechende Fusionsprotein kann aus dem Periplasma (Hellebust *et al.*, 1998) oder aus dem Kulturmedium gewonnen werden (Mergulhao *et al.*, 2004; Moks, et al., 1987).

Die Bildung von Disulfidbrücken im Periplasma wird durch Disulfid Oxidoreduktasen bzw. Isomerasen katalysiert und kontrolliert (Raina und Missiakas, 1997; Rietsch und Beckwith, 1998), deren wichtigste Komponente ist die Disulfid Oxidoreductase DsbA (20 kDa). Auch DsbA selbst kann als N-terminaler Fusionspartner zur periplasmatischen Proteinproduktion eingesetzt werden und die Ausbeuten verbessern (Collins-Racie *et al.*, 1995; Winter *et al.*, 2000).

Barnase (12 kDa), eine enzymatisch inaktive Variante der extrazellulären RNAse aus *Bacillus amyloliquefaciens* wird bei Expression in *Escherichia coli* ins Periplasma und ins Nährmedium sezerniert und kann als Fusionsdomäne für die native, sekretorische Produktion von Cystin-Knoten Peptiden eingesetzt werden (Schmoldt *et al.*, 2005).

Nachteilig für alle bisher bekannten Systeme ist, dass sich viele Proteine mit den verfügbaren Fusionsproteinen nur unzureichend produzieren lassen. Der Erfindung lag somit die Aufgabe zu Grunde ein neues periplasmatisches Fusionsprotein zu entwickeln, welches mit guter Ausbeute exprimiert und mit hoher Effizienz sekretiert, wobei die Eigenschaften des gewählten N-terminalen Fusionspartners entscheidend sind. Erforderlich ist, dass diese N-terminale Domäne über eine hohe Stabilität verfügt und leicht zur nativen Struktur faltet. Ihre Struktur darf die Faltung des Zielproteins nicht behindern, sondern soll diese fördern. Darüber hinaus ist dieser Teil des Fusionsproteins möglichst klein zu halten, so dass es möglich ist, auch größere Zielproteine als Fusionen zu exprimieren.

Erfindungsgemäß wurde das Problem dadurch gelöst, dass ein Fusionsprotein entwickelt wurde, dessen N-terminale aus Ecotin besteht. Ecotin ist ein periplasmatisches Protein aus *Escherichia coli* (Chung *et al.*, 1983), und ein potenter Inhibitor einer Vielzahl von Serin-Proteasen. Es bildet durch mehrere zentrale antiparallele ß-Faltblätter eine sehr stabile Tertiärtruktur aus (Fig. 1). Die primären und sekundären Wechselwirkungen mit Trypsin oder anderen Proteasen erfolgen an zwei benachbarten Loopstrukturen auf der Oberfläche. Das Protein dimerisiert durch "domain swapping" der C-terminale Domänen. Die freien C-Termini ragen aus dem Dimer in entgegengesetzter Richtung heraus (McGrath *et al.*, 1995). Durch Mutagenese gezielter Oberflächenreste wurden aus Ecotin bereits Protease-Inhibitoren mit veränderter Spezifität entwickelt und selektiert (Stoop und Craik, 2003; Wang *et al.*, 1995).

Gemäß Anspruch 1 wird mittels bekannter molekularbiologischer Techniken ein Fusionsprotein mit der Aminosäuresequenz für die Ecotin-Signalsequenz, für Ecotin, und das rekombinante Protein konstruiert. Eine vorteilhafte Ausführung stellt die Generierung einer kurzen Peptidsequenz, beispielsweise eine Folge von sechs Glycin-Serin Resten, zwischen Ecotin und dem rekombinanten Protein (vgl. Fig. 2A) dar. Eine weitere beinhaltet die Verbindung beider Proteine durch eine Peptidsequenz, die eine gezielte Abspaltung mittels einer Protease ermöglicht. Hier sind inhibitorisch inaktive Ecotin-Varianten vorteilhaft. Eine weitere Ausführung stellt ein entsprechendes monomeres Fusionsprotein dar, das bevorzugt durch eine Insertion von Ala-Asp-Gly nach Trp130 (Eggers *et al.*, 2001) oder auch durch C-terminale Verkürzung des Proteins um 10 Aminosäuren (Pál *et al.*, 1996) hergestellt wird.

Das erfindungsgemäße Verfahren wird zur Herstellung therapeutisch relevanter disulfidverbrückter Proteine, u. a. von Aspartat-Proteasen, eingesetzt. Aspartat-Proteasen haben Schlüsselfunktionen bei der Entstehung vieler Krankheiten. So ist bekannt, dass Aspartat-Proteasen an der Prozessierung des Alzheimer Peptids (Sinha *et al.*, 1999; Vassar *et al.*, 1999) an der Kontrolle von Blutdruck und Elektrolythaushalt (Suzuki *et al.*, 2004), an der Prozessierung des HIV-1 gag Polyproteins zu den viralen Matrix- und Strukturproteinen (Erickson-Viitanen *et al.*, 1989) beteiligt sind. Lysosomale und endosomale Aspartat-Proteasen kontrollieren den zellulären Proteinstoffwechsel und sind an Präsentation von Antigenen auf der Zelloberfläche beteiligt (Conner, 2004; Hewitt *et al.*, 1997; Maric *et al.*, 1994). So wird Cathepsin D als prognostischer Marker bei der Tumorentstehung eingesetzt (Scorilas *et al.*, 1999). Pepsin und seine Proform dienen als Serum-Marker für die Diagnostik schwerer atrophischer Gastritis und Magenkrebs.

Aspartat-Proteasen sind sezernierte Proteine, die in den eukaryontischen Zellen als Proproteine synthetisiert und später zur aktiven Protease prozessiert werden. Sie enthalten mehrere konservierte Disulfidbrücken. Mit Ausnahme der HIV Protease fallen alle bislang in

Bakterien produzierten Aspartat-Proteasen intrazellulär als *inclusion bodies* aus. Ein mögliches Renaturierungsverfahren unter alkalischen Bedingungen wird in US Patent 6,583,268 beschrieben (Lin, 2003). Es verläuft jedoch nicht für alle Aspartat-Proteasen erfolgreich. Das erfindungsgemäße Verfahren der nativen Expression im Periplasma stellt somit eine echte Alternative dar.

Unter den Aspartat-Proteasen sind Pepsin und seine Proform Pepsinogen am besten untersucht. Ziel war daher, dieses Protein als Modellprotein für Aspartat-Proteasen in seiner natürlichen disulfidverbrückten Form im bakteriellen Periplasma zu produzieren. Bei Verwendung kurzer bakterieller Signalpeptide konnte keine Sekretion von Pepsinogen ins Periplasma nachgewiesen werden, statt dessen kam es zur Bildung von *inclusion bodies*. Durch Fusion von Pepsinogen an das periplasmatische Protein Ecotin dagegen lässt sich natives Pepsinogen im Periplasma nachweisen. Das resultierende Fusionsprotein faltet im Periplasma zur natürlichen, disulfidhaltigen Struktur, und kann in dieser Form aus dem Periplasma aufgereinigt werden. Aus dem Fusionsprotein kann die Protease Pepsin nach Aktivierung autokatalytisch gespalten werden.

Des Weiteren wurde ein Verfahren zur Isolierung von nativem disulfidverbrückten Proteinen aus dem Periplasma entwickelt. Ein Problem ist, dass das rekombinante Protein während der Produktion im Periplasma aber auch bei der Aufarbeitung oxidativ missfalten kann, d.h. es entstehen intramolekulare oder intermolekulare unnatürliche Disulfidbrücken, die die Gesamtausbeute verringern. Aus USP 6,455,279 ist bekannt, dass die Ausbeute an natürlich gefaltetem, disulfidverknüpften Protein durch den Zusatz niedermolekularer Thiolreagenzien ins Nährmedium erhöht werden kann (Glockshuber *et al.*, 2001).

Die Präparation von bakteriellem Periplasma erfolgt in der Regel durch osmotischen Schock. Die Bakterienzellen werden kurzzeitig in einer Pufferlösung mit hohem osmotischen Wert inkubiert und anschließend in ein Medium niedriger Osmolarität, z.B. Wasser, verdünnt.

Erfindungsgemäß wurde ein Verfahren entwickelt, in welchem die Bildung unnatürlicher Disulfidbrücken während des osmotischen Schocks durch den Zusatz niedermolekularer Thiole unterdrückt wird. Während des osmotischen Schocks kommt es (insbesondere bei der Verarbeitung großer oder gefrorener Zellmassen) immer auch zur partiellen Lyse. Cytoplasmatische Proteine werden freigesetzt und bilden unnatürliche Disulfid-Heterodimere mit den periplasmatischen Proteinen. In dem erfindungsgemäßen Aufschlussverfahren wird durch den Zusatz von 1 mM oxidiertem und reduziertem Glutathion die Bildung unnatürlicher intra- und intermolekularer Disulfidbrücken wirkungsvoll unterdrückt. So kann in drei Reinigungsschritten homogenes Protein gewonnen werden (Fig. 4A) während der Ansatz ohne Glutathion inhomogenes Produkt liefert (Fig. 4B).

Eine Fusion aus Ecotin und humanem Proinsulin dient als ein weiteres Beispiel für die native periplasmatische Produktion eines therapeutisch bedeutenden, disulfidverbrückten Proteins. Proinsulin ist die biologische Vorstufe für Insulin, ein Peptidhormon mit blutzuckersenkender Wirkung. Natives Proinsulin besteht aus 86 Aminosäuren mit drei Disulfidbrücken (Fig. 6). Bei der Prozessierung zu Insulin wird aus Proinsulin ein Polypeptid, das sog. C-Peptid (connecting peptide) herausgespalten. Die verbleibenden N-und C-terminalen Polypeptide werden als B- und A-Kette von Insulin bezeichnet. Sie enthalten die in der Proform gebildeten Disulfidbrücken, wobei zwei der Disulfidbrücken die A- und B-Ketten intermolekular miteinander verbinden, während die dritte Disulfidbrücke zwei Cysteine der A-Kette intramolekular miteinander verknüpft. Proinsulin wird in den ß-Zellen der Langerhansschen Inseln der Pankreas synthetisiert, wo in den sekretorischen Granula durch Furin/Prohormonconvertasen diese Prozessierung zum Insulin stattfindet.

Die Disulfidbrücken im Proinsulin sind nicht linear angeordnet (Fig. 6). Die rekombinante Herstellung im Periplasma von *E. coli* in nativer Form ist schwierig, da das Protein leicht aggregiert, nur unvollständig sezerniert und durch Proteasen abgebaut wird (Kang und Yoon, 1994; Mergulhao, et al., 2004; Schäffner *et al.*, 2001; Winter, et al., 2000).

Erfindungsgemäß wird Proinsulin mit Ecotin fusioniert (Beispiel 5). Nach Expression in *E. coli* BL21(DE3) ist das Fusionsprotein in der löslichen Periplasmafraktion. Ein Hexahistidin-tag zwischen beiden Proteindomänen kann zur Aufreinigung genutzt werden. Nach Periplasma-Aufschluss und Ni-Affinitätschromatographie wird mit Thrombin gespalten. Die Thrombin-Schnittstelle liegt in der Peptidlinker-Region zwischen beiden Proteindomänen (Fig. 7A) und führt zur Abspaltung von Ecotin. Im Gegensatz zu anderen Serin-Proteasen wird Thrombin durch Ecotin nur geringfügig inhibiert. Natives Proinsulin (korrekte Disulfidverbrückung) wird nachfolgend im ELISA quantifiziert, die Identität durch HPLC und Massenspektrometrie bestätigt.

Beispiel 1

Konstruktion eines Ecotin-Pepsinogen Fusionsproteins

Um humanes Pepsinogen an Ecotin zu fusionieren, wird das Gen für *E. coli* Ecotin einschließlich seiner Signalsequenz (GenBank M60876) aus chromosomaler DNA von *E. coli* JM83 mit den Oligonucleotiden Ecotin-Fwd (5'-TTC TAA GAA TTC GAA GGA GAT ATA CAT AAT GAA GAC CAT TCT ACC TGC A-3') und Ecotin-GS-Rev (5'-ATC TTA TCC GGA ACC AGA ACC AGA ACC ACG AAC TAC CGC ATT GTC AAT TT-3') amplifiziert. Humanes Pepsinogen A wird aus pHQPEX-30-5 mit den Oligonucleotiden GS-Peps-Fwd (5'-GTT ATA TCC GGA TCT GGT TCT GGT TCT GGT TAC AAG GTC CCC CTC A-3') und Peps-His-Rev (5'-GTA TAT GTC GAC TTA GTG GTG GTG GTG GTG

GTG AGC CAC GGG GGC CAG ACC-3') amplifiziert. Beide Fragmente werden in pCR-Blunt II-TOPO kloniert. Nach Restriktionsspaltung mit *Eco*RI und *Bsp*E1 (Ecotin-Fragment) bzw. *Bsp*E1 und *Sal*I (Pepsinogen-Fragment) entsteht durch Ligation beider PCR-Fragmente das Gen für ein Fusionsprotein, bei dem Ecotin und Pepsinogen über 12 Glycin-Serin-Reste in Folge miteinander verbunden sind (Fig. 2A). Das Gen wird über *Eco*RI und *Bsp*E1 in den Expressionsvektor pTrc99a einkloniert. Das resultierende Plasmid wird mit pEGP1 bezeichnet.

Beispiel 2

Expression des Ecotin-Pepsinogen Fusionsproteins

Das Plasmid pEGP1 wird in *E. coli* BL21(DE3) transformiert. Aus einer Übernacht-Kultur bei 37 °C wird 2 % in 20 ml LB Medium mit 100 mg L^{-1} Ampicillin überimpft (37 °C, 140 rpm). Die Induktion der Produktion des Fusionsproteins erfolgt bei einer Optischen Dichte von 0,5 – 0,7 durch Zusatz von 1 mM IPTG. Eine Inkubationszeit von 4 h bei 24 °C eignet sich für die Produktion von nativem Fusionsprotein bevorzugt.

Für die Analyse des Periplasmas wird ein Aufschluss durch osmotischen Schock durchgeführt Pro ml Kultur einer optischen Dichte von 1.0 wird ein Periplasma-Extrakt mit einem Volumen von 100 µl in 0,5 mM $MgCL_2$, 1 mM EDTA, 10 mM Tris-HCl pH 8,0 erhalten.

Der Nachweis von Pepsinogen im Periplasma-Extrakt erfolgt durch Bestimmung der proteolytischen Aktivität. Aus der Ecotin-Pepsinogen-Fusion wird unter sauren Bedingungen Pepsin abgespalten, dessen proteolytische Aktivität anhand der Spaltung von enhanced green fluorescent protein (EGFP) bestimmt werden kann (Malik *et al.*, 2005). Zur Prozessierung von Pepsinogen wird der Extrakt zunächst auf 37 °C temperiert, mit 0,1 Volumen 1 M Citrat pH 2,0 versetzt und 10 min inkubiert. Die Proteolyse von säuredenaturiertem EGFP (11 µg)

erfolgt mit 50 µl Protease bzw. Periplasma-Extrakt. Unter den oben beschriebenen Anzuchtbedingungen wird pro mL Bakterienkultur einer Optischen Dichte von 1,0 wurden ca. 100 ng natives Fusionsprotein (70 ng Pepsinogen) erhalten.

Für Produktion des Fusionsproteins im Fermentor wird ein Komplexmedium (0,5 % Glucose, 5 % Hefeextrakt, 1,1 % KH_2PO_4, 0,05 % NH_4Cl, 0,068 % $MgCl_2$) verwendet. Das Kulturvolumen beträgt 8 L. Die Fermentation wird zunächst bei 37 °C und pH 7,0 geführt, bis zum vollständigen Glucose-Verbrauch (Optische Dichte 18,7). Dann wird die Produktion des Fusionsproteins durch Zugabe von 1 mM IPTG induziert. Die Temperatur wird auf 24 °C reduziert. Während der weiteren Fermentation für 5 h wird eine Feeding-Lösung aus 30 % Hefeextrakt und 25 % Glycerin mit einer kontinuierlich ansteigenden Rate von 1.5 – 4,5 ml min^{-1} zugesetzt. Nach 5 h werden die Zellen durch Zentrifugation sedimentiert. Die erhaltene Biomasse beträgt ca. 600 g (OD 45).

Beispiel 3

Periplasma-Aufschluss und Reinigung des Ecotin-Pepsinogen Fusionsproteins

30 g gefrorene Biomasse werden mit 30 ml einer Pufferlösung aus 50 % Saccharose, 20 mM Tris-HCl, 10 mM EDTA, pH 9,0 vollständig resuspendiert. Die Zellmasse wird für 2 h bei 4 °C langsam gerührt (500 rpm) (Rathore *et al.*, 2003). Anschließend erfolgt ein osmotischer Schock durch Zugabe von 300 ml dest. Wasser (0 °C), und weiteres Rühren bei 4 °C für 2 h. Protoplasten und unlösliche Zellbestandteile werden nachfolgend durch Zentrifugation 15 min, 10.000 g 4°C, abgetrennt.

Das bekannte Verfahren wird wie folgt modifiziert:

1.) Um eine Aggregation und Missfaltung des rekombinanten Proteins während des Periplasma-Aufschlusses zu verhindern, werden im hier beschriebenen Aufschlussverfahren niedermolekulare Thiole, vorzugsweise Glutathion zugesetzt: Die o.g. Pufferlösung und das destillierte Wasser werden jeweils 1 mM oxidiertem und reduziertem Glutathion supplementiert.

2.) Nach osmotischem Schock und Zentrifugation wird der Überstand mit 6 mM $MgCl_2$ versetzt und mit 6µl (500 U/µl) Benzonase (Merck) für 45 min bei 25 °C inkubiert.

3.) Für die nachfolgende Ni-NTA Chromatographie wird der Periplasma-Extrakt bei 4°C gegen 50 mM Tris-HCl, 300 mM NaCl pH 8,0 dialysiert. Nach Zusatz von 10 mM Imidazol werden unlösliche Bestandteile abzentrifugiert (10.000 g 20 min). Der Überstand (ca. 360 ml) wird durch eine 0,45 µm Membran filtriert.

Der Periplasmaextrakt wird mit 40 ml äquilibriertem His-Bind Gel (Novagen) versetzt. Die Suspension wird bei 4 ° für 4 h langsam gerührt (100 rpm). Anschließend wird das sedimentierte Gelmaterial mit drei Volumina 50 mM Tris-HCl, 300 mM NaCl, 10 mM Imidazol, pH 8,0 gewaschen. Die Elution des Fusionsproteins erfolgt durch 3 Volumina 50 mM Tris-HCl, 300 mM NaCl, 250 mM Imidazol, pH 8,0.

Die eluierten Fraktionen werden vereinigt, gegen 50 mM Tris-HCl pH 7,5 dialysiert, und zentrifugiert (20.000 g, 20 min, 4°C). Der Überstand wird filtriert (0,2 µm). Je 50 ml werden mit einer Flussrate von 1 ml min^{-1} auf 5 ml HiTrapQ HP (Amersham) gegeben. Die Säule wird mit 70 ml des o.g. Puffers gewaschen. Die Elution erfolgt durch einen zweistufigen Gradienten (45 ml 0-30 % NaCl mit 1,5 ml min^{-1}, und 5 ml 30 -100 % NaCl mit 1 ml min^{-1}) im o.g. Puffer.

Fig. 3 zeigt Proben der Aufreinigung. Das Fusionsprotein (58 kDa) ist bereits im Periplasma-Extrakt (Spur 1) erkennbar. Wird der Extrakt mit 0,1 Vol Citrat pH 2,0 versetzt, erfolgt Proteolyse (Spur 9). Proben nach Ni-NTA Chromatographie (Spur 3 und 4) zeigen eine deutliche Anreicherung des Fusionsproteins. Während der anschließenden Dialyse verbleiben ca. 70 % im Überstand, während 30 % als Aggregat ausfallen. Nach Anionenaustausch-Chromatographie and HiTrapQ ist die Ecotin-Pepsinogen-Fusion bereits deutlich angereichert (ca. 73 % homogen, Spur 8).

Nach HiTrapQ Chromatographie werden entsprechende Fraktionen vereinigt und durch Gelfiltration an Superdex G-75 (Amersham) in Gegenwart von 50 mM Tris-HCl, 300 mM NaCl, pH 7,5 bei einer Flussrate von 2 ml min^{-1} aufgetrennt. Fig. 4 zeigt Fraktionen der abschließenden Gelfiltration. In Fig. 3A wurde als Ausgangsmaterial für die Reinigung ein Periplasma-Aufschluss in Gegenwart von 1 mM oxidiertem und reduziertem Glutathion (s.o.) durchgeführt. In Fig. 3B wurde ein identisches Reinigungsprotokoll verwendet, Ausgangsmaterial war jedoch ein Periplasma-Aufschluss in Abwesenheit von Thiolen. In Gegenwart von Thiolen (Fig. 4A) liegt nach der Gelfiltration homogenes Fusionsprotein vor, in Abwesenheit von Thiolen (Fig. 4B) wird das cytoplasmatische SlyD Protein (21 kDa) mit aufgereinigt.

Tabelle 1 zeigt das Reinigungsprotokoll in der Übersicht. Nach Aufschluss von 30 g Biomasse enthält der Periplasma-Extrakt ca. 600 mg Gesamtprotein. Nach Dialyse und Ni-NTA-Chromatographie verbleiben 47 mg, nach HiTrapQ 11,5 mg. Das Fusionsprotein liegt nun bereits in nahezu homogener Form vor (vgl. Abb. 2, Spur 8). Durch abschließende Gelfiltration werden 5,2 mg vollständig homogenes Präparat erhalten. Dies entspricht einer Ausbeute von 22,9 %.

Beispiel 4

Prozessierung des gereinigten Ecotin-Pepsinogen Fusionsproteins zu Pepsin

Die Prozessierung des Ecotin-Pepsinogen-Fusionsproteins zu Pepsin erfolgt säurekatalysiert. Das Fusionsprotein (30 µg/ml) wird mit 0,1 Vol. 1 M Citrat pH 2,0 versetzt und bei 37 °C inkubiert. Fig. 5 zeigt Proben der Prozessierung nach unterschiedlicher Inkubationszeit. Die N-terminale Domäne (Ecotin) ist bereits nach 30 Sekunden vollständig abgebaut, die entstehenden Spaltprodukte entsprechen Pepsinogen (40 kDa) und Pepsin (34 kDa). Erst nachfolgend wird das entstandene Pepsinogen vollständig zu Pepsin prozessiert.

Beispiel 5

Konstruktion des Ecotin-Proinsulin Fusionsproteins

Humanes Proinsulin wird aus pDSBA3-PI (Winter, et al., 2000) mit den Oligonucleotiden PI-Fwd (5‘-GGT TCC GAT CTG GTT CTG GTT CTC TGG TCC CCC GCG GTA GTC ACC ACC ACC ACC ACC ACC GTT TTG TGA ACC AAC ACC TGT GCG GC-3‘) und PI-Rev (5’- AGT GTC GAC TTA GTT GCA GTA GTT CTC CAG CTG GTA-3’) amplifiziert und in pCR-Blunt II-TOPO kloniert. Nach Restriktionsspaltung mit *Bsp*E1 und *Sal*I wird das Pepsinogen-Fragment in pEGP1 durch das entsprechende Proinsulin-Fragment ersetzt. Das resultierende Plasmid wird mit pEGPI1 bezeichnet.

Beispiel 6

Produktion des Ecotin-Proinsulin Fusionsproteins

Das Plasmid pEGPI1 wird in *E. coli* BL21(DE3) transformiert. Aus einer Übernacht-Kultur bei 37 °C wird 1 % in 1 L LB Medium mit 100 mg L^{-1} Ampicillin überimpft (37 °C, 140 rpm). Die Induktion der Produktion des Fusionsproteins erfolgt bei einer Optischen Dichte von 0,5 – 0,7 durch Zusatz von 1 mM IPTG. Eine Inkubationszeit von 4 h bei 24 °C eignet sich für die Produktion bevorzugt. Aus 4 L Kultur werden ca. 15 g Biomasse erhalten.

Der Periplasma-Aufschluss erfolgt wie in Beispiel 3 beschrieben. Auf den Zusatz von Glutathion wird verzichtet. Die Vorbereitung der Ni-Affinitätschromatographie erfolgt wie in Beispiel 3 beschrieben. Die Chromatographie wird an 15 ml Ni-NTA Agarose in 50 mM Tris, 10 mM Imidazol, pH 8,0 durchgeführt bei einer Flussrate von 1 ml/min. Die Elution erfolgt mit 0 – 100 % NaCl. Der Nachweis von Proinsulin erfolgt durch ELISA (Winter, et al., 2000). Die entsprechenden Fraktionen werden vereinigt und gegen 50 mM Tris, 300 mM NaCl pH 8,0 dialysiert. Ecotin ist ein dimeres Protein, so dass auch das Fusionsprotein zunächst als Homodimer vorliegt. Die Abspaltung von Ecotin erfolgt durch Thrombin. Es wird monomeres Proinsulin frei. Die Spaltung erfolgt bei Raumtemperatur für 2 h. Nach Zusatz von PMSF und 10 mM Imidazol wird 30 min bei 4°C und 25000 Upm zentrifugiert.

Der Hexahistidin-tag Er befindet sich nach Spaltung mit Thrombin im N-Terminus von Proinsulin. Er kann nun genutzt werden, um Proinsulin und Ecotin zu trennen. Die Proteinlösung wird erneut auf die Ni-NTA-Agarose aufgetragen, die Elution von Proinsulin erfolgt in 250 mM Imidazol. Die Fraktionen werden im ELISA analysiert. Das spezifische Signal im ELISA ist nun nahezu verdoppelt. Es liegt natives monomeres Proinsulin vor. Die Identität wird durch HPLC und Massenspektrometrie bestätigt.

Patentansprüche

1. Fusionsprotein zur Produktion von rekombinanten Proteinen oder Peptiden im Periplasma enthaltend N-terminal die Aminosäuresequenz für die Ecotin-Signalsequenz und für Ecotin sowie C-terminal für das rekombinante Protein oder Peptid.

2. Fusionsprotein nach Anspruch 1, gekennzeichnet dadurch, dass eine Peptidsequenz zwischen Ecotin und dem rekombinantem Protein oder Peptid generiert wird.

3. Fusionsprotein nach Anspruch 1, gekennzeichnet dadurch, dass die Peptidsequenz zwischen Ecotin und dem rekombinanten Protein oder Peptid eine Protease-Schnittstelle enthält.

4. Fusionsprotein nach Anspruch 1 bis 3, gekennzeichnet dadurch, dass eine inhibitorisch inaktive Ecotin-Variante verwendet werden kann.

5. Fusionsprotein nach Anspruch 1 bis 4, gekennzeichnet dadurch, dass eine monomere Ecotin-Variante eingesetzt werden kann.

6. Verfahren zur Herstellung therapeutisch relevanter disulfidverbrückter Proteine oder Peptide gekennzeichnet dadurch, dass die Proteine oder Peptide mittels Fusion an *Escherichia coli* Ecotin und anschließender nativer Expression im Periplasma gewonnen werden.

7.Verfahren nach Anspruch 6 gekennzeichnet dadurch, dass während der Präparation des Periplasmas niedermolekulare Thiole zugesetzt werden.

8. Verfahren nach Anspruch 6, gekennzeichnet dadurch, dass als Thiol 1mM oxidiertes und reduziertes Gluthation eingesetzt wird.

9. Verfahren nach Anspruch 6, gekennzeichnet dadurch, dass als rekombinantes disulfidverbrücktes Protein eine Aspartat-Protease verwendet werden kann.

10. Verfahren nach Anspruch 6, gekennzeichnet dadurch, dass als rekombinantes disulfidverbrücktes Protein humanes Proinsulin eingesetzt werden kann.

Tabelle 1

Reinigung des Ecotin-pepsinogen Fusionsproteins

Schritt	Vol (ml)	Gesamtprotein (mg)[a]	Fusionsprotein (mg) [c]	Ausbeute (%)
Periplasma-Aufschluss	360	612,6	8,9	100,0
Ni-NTA Pool, dialysiert	190	47,0	4,8	54,8
HiTrapQ Chromatographie	9	11,5	4,1	46,8
Sephadex G-75 Chromatographie	22	5,2[b]	2,0	22,9

[a] Das Gesamtprotein wurde nach Bradford bestimmt.

[b] Das gereinigte Protein wurde mit A_{280} bestimmt, mit $\varepsilon = 83240 \times Mol^{-1} \times cm^{-1}$ (Ecotin-Pepsinogen Fusion)

[c] Natives Fusionsprotein wurde bestimmt über die proteolytische Aktivität, als Referenz diente Pepsinogen aus Schwein (Sigma).

1 mg Pepsin entspricht 1,16 mg Pepsinogen bzw. 1,67 mg Fusionsprotein.

Fig. 1: Ecotin (*Eschericha coli* Trypsin Inhibitor, 16 kDa)

Dargestellt ist die Struktur des dimeren Proteins. Die C-terminalen Domänen beider Untereinheiten bilden zusammen ein antiparalleles ß-Faltblatt, die C-termini (C bzw. C´) zeigen in entgegengesetzte Richtungen. Die letzten 4 Aminosäuren (139-142) sind rechtwinklig zum ß-Faltblatt angeordnet. Die Bindung von Serin-Proteasen erfolgt an den flexiblen Loopstrukturen, die in der Abb. nach oben und unten aus dem Dimer herausragen. Die Abbildung wurde mit Pymol (http://pymol.sourceforge.net/) mit 1ECZ aus der PDB Proteindatenbank erstellt.

Fig. 2: Ecotin-Pepsinogen Fusionsprotein

Die Aminosäuren 1-20 entsprechen der Ecotin-Signalsequenz. Zwischen Ecotin (Aminosäuren 21-162) und Pepsinogen (Aminosäuren 174-546) wurde ein Peptidlinker aus 6 x Glycin-Serin eingefügt. Für die Reinigung an Ni-NTA wurden am C-Terminus 6 Histidinreste eingefügt (Aminosäuren 547-552). Der Pfeil zeigt die Spaltstelle für die Prozessierung von Pepsinogen zu Pepsin.

Fig. 3: SDS-PAGE Analyse bei Expression und Reinigung des Ecotin-Pepsinogen Fusionsproteins

Der Pfeil links zeigt das Ecotin-Pepsinogen Fusionsprotein (58 kDa). Aufgetragen wurden in Spur 1, Proteingrössenstandard (Peqlab); 2, Periplasma-Extrakt; 3-5 Ni-NTA Chromatographie, 6, Überstand nach Dialyse, 7, Aggregat nach Dialyse, 8, HiTrapQ

Chromatographie, 9, ein Periplasma-Extrakt nach Aktivierung von Pepsinogen zu Pepsin (Zusatz von 0,1 Vol 1 M Citrat pH 2,0).

Fig. 4: Effekt von Glutathion auf die Reinigung von Ecotin-Pepsinogen aus dem Periplasma

Dargestellt sind SDS PAGE-Analysen nach Reinigung von Ecotin-Pepsinogen aus dem Periplasma (Beispiel 3). Beide Gele zeigen den letzten Reinigungsschritt, Fraktionen nach Gelfiltration an Superdex G-75. In A wurde dem Periplasma-Extrakt 1mM oxidiertes und reduziertes Glutathion zugesetzt, in B nicht. Der obere Pfeil zeigt das Ecotin-Pepsinogen Fusionsprotein (58 kDa), der untere Pfeil SlyD (20 kDa) aus *E. coli*, das unter den Bedingungen in B coeluiert.

Fig. 5: Säurekatalysierte Aktivierung des gereinigten Ecotin-Pepsinogen-Fusionsproteins

Das gereinigte Ecotin-Pepsinogen Fusionsprotein wird nach Zusatz von Säure autokatalytisch zu Pepsin prozessiert (Beispiel 4). Freigesetztes Pepsin spaltet zunächst die Ecotin-Domäne, dann wird auch restliches Pepsinogen vollständig zu Pepsin prozessiert. Spur 1, Proteinstandard (Peqlab); Spur 2, Ecotin-Pepsinogen-Fusionsprotein nicht aktiviert; Spuren 3-10, Zusatz von 0,1 Vol. 1 M Citrat pH 2,0: 30s, 1 min, 2 min, 4 min, 6 min, 8 min, 10 min und 16 min. Der obere Pfeil zeigt das Fusionsprotein, der mittlere Pfeil Pepsinogen, der untere Pepsin.

Fig. 6: Humanes Proinsulin

Gezeigt ist eine Abbildung nach (Mackin *et al.*, 1998). Die beiden intermolekularen Disulfidbrücken der A- und B-Kette und die intramolekulare Disulfidbrücke der A-Kette sind mit Linien gezeigt. Die natürlichen Spaltstellen für die Prozessierung von Proinsulin zu Insulin durch die Proteasen PC1, PC2 und CPE sind durch Pfeile gekennzeichnet.

Fig. 7: Ecotin-Proinsulin Fusionsprotein

Die Aminosäuren 1-20 entsprechen der Ecotin-Signalsequenz. Zwischen Ecotin (Aminosäuren 21-162) und Proinsulin (189-273) befindet sich ein Peptidlinker aus 6 x Glycin-Serin, gefolgt von der Sequenz für eine Thrombin-Spaltstelle für die Spaltung des Fusionsproteins in Ecotin und Proinsulin. Der N-terminus von Proinsulin wurde mit einem Affinitäts-tag aus 6 Histidin-Resten sowie einem Arginin zur Prozessierung durch Trypsin versehen.

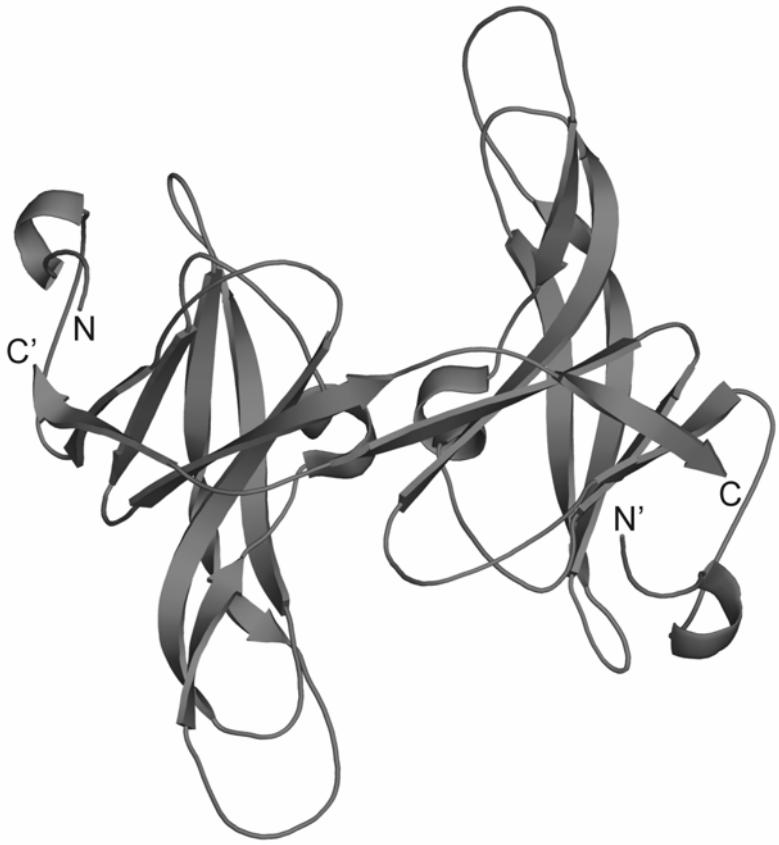


Fig. 1

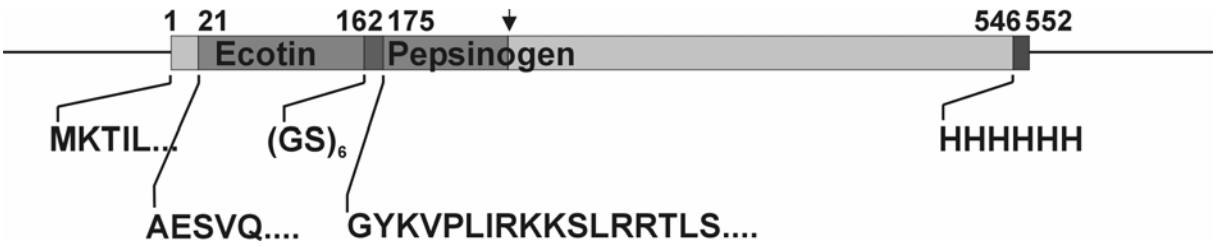

1 21
162 175
546 552
Ecotin
Pepsinogen
MKTIL...
(GS)6
HHHHHH
AESVQ....
GYKVPLIRKKSLRRTLS....

Fig. 2

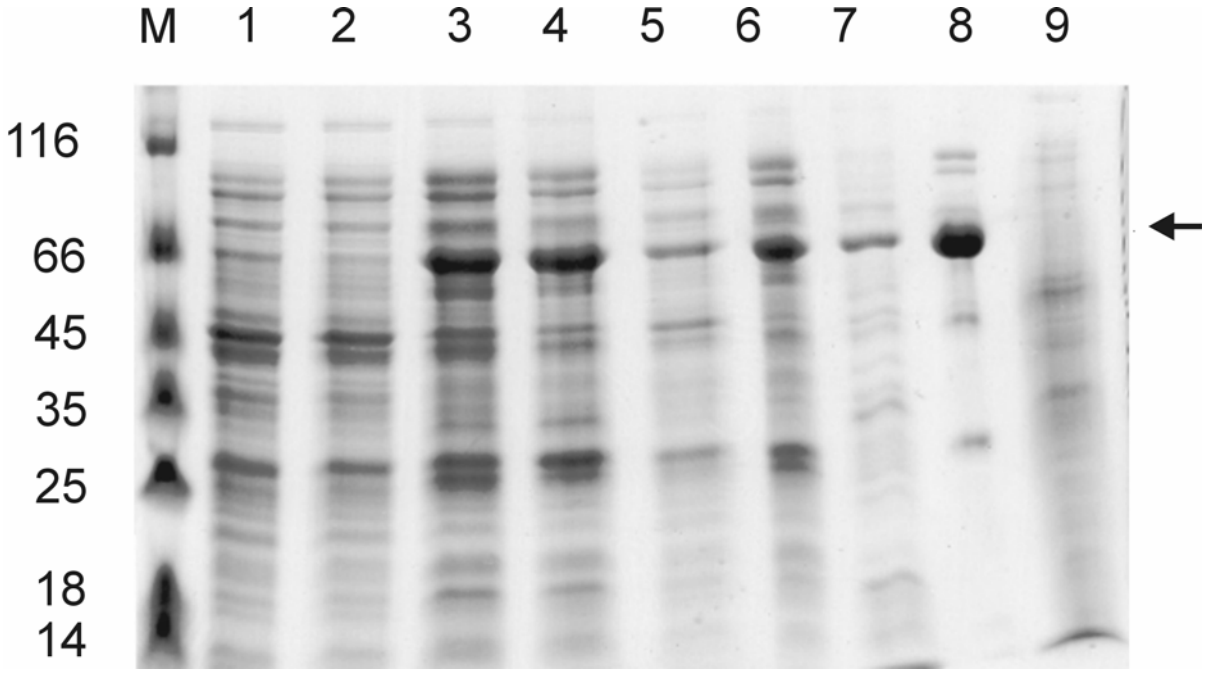


Fig. 3

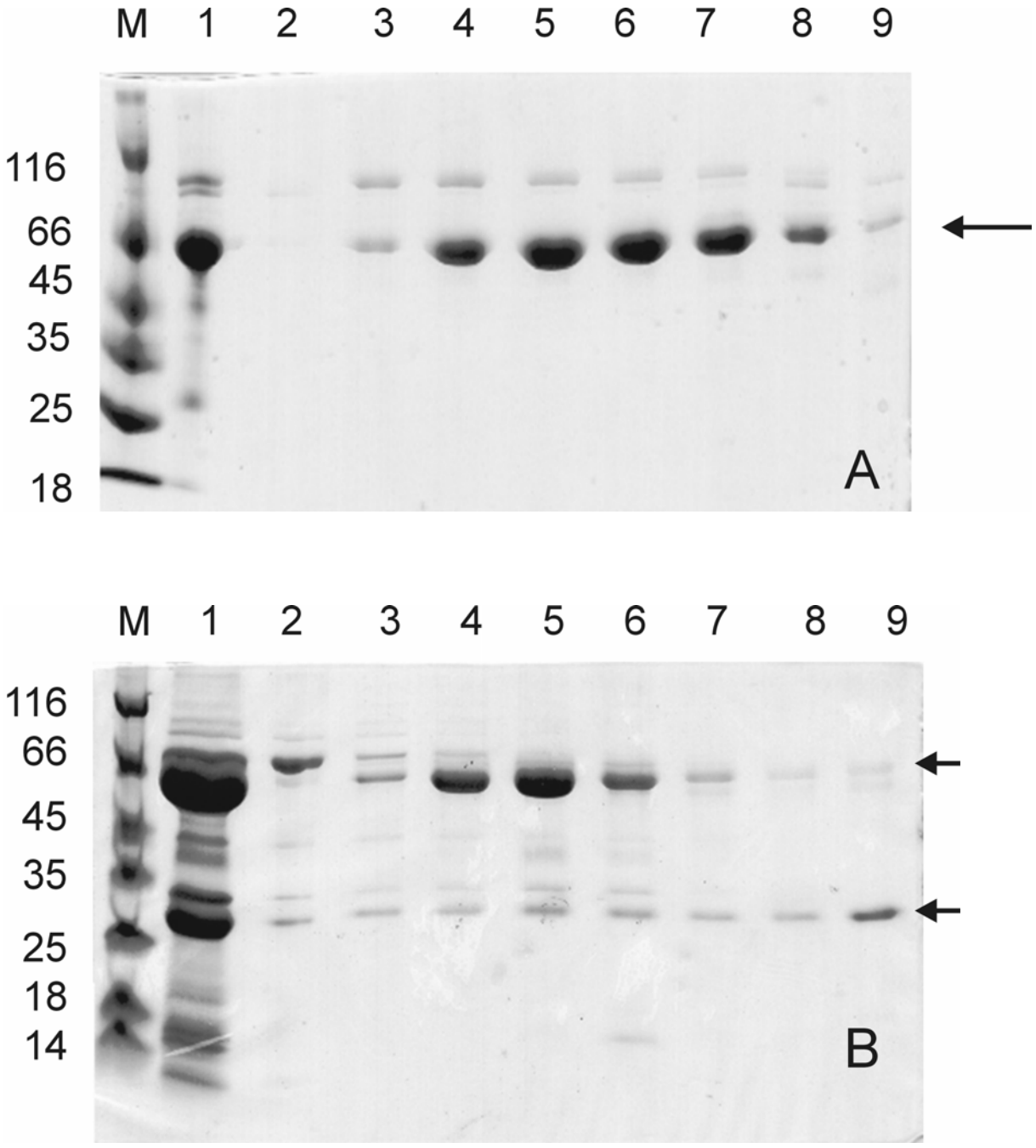

M 1 2 3 4 5 6 7 8 9
116
66
45
35
25
18
A
M 1 2 3 4 5 6 7 8 9
116
66
45
35
25
18
14
B

Fig. 4

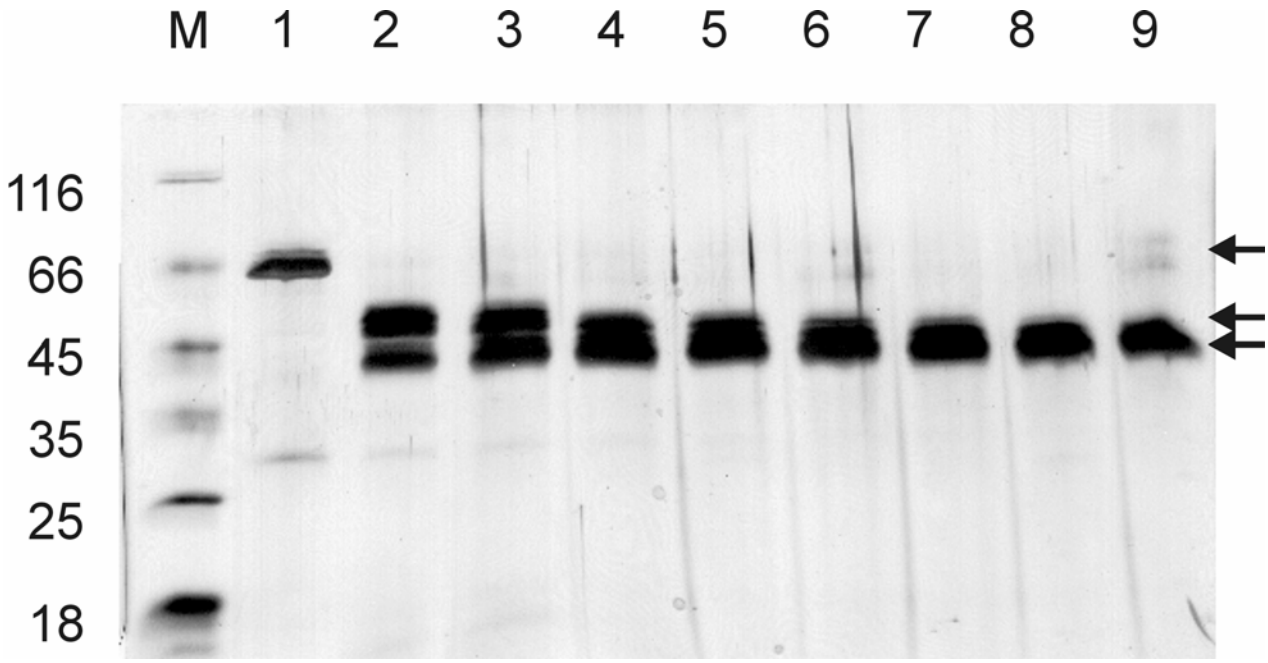


Fig. 5

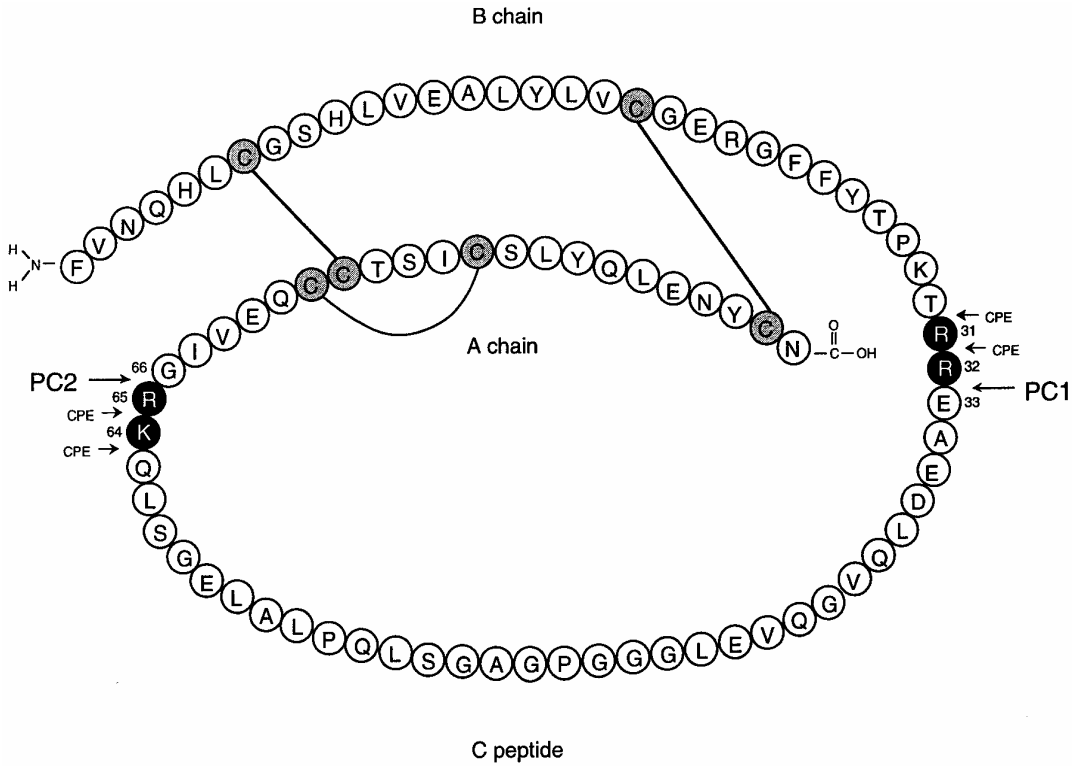

B chain
A chain
C peptide
PC2
PC1
CPE
CPE
CPE
CPE
66
65
64
31
32
33
OH

Fig. 6

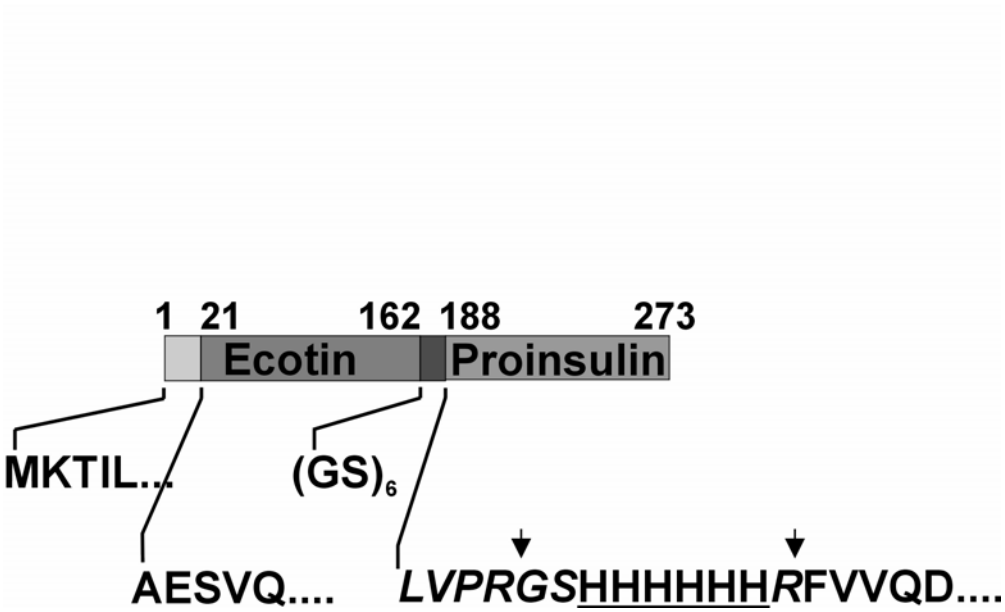

1 21 162 188 273
Ecotin
Proinsulin
MKTIL...
(GS)6
AESVQ....
LVPRGSHHHHHHRFVVQD....

Fig. 7

Literatur:

Chung, C.H., Ives, H.E., Almeda, S., Goldberg, A.L., Purification from *Escherichia coli* of a periplasmic protein that is a potent inhibitor of pancreatic proteases, J. Biol. Chem. 258 (1983) 11032-11038.

Collins-Racie, L.A., Mc Colgan, J.M., Grant, K.L., Di Blasio-Smith, E.A., Mc Coy, J.M., La Vallie, E.M., Production of recombinant bovine enterokinase catalytic subunit in *Escherichia coli* using the novel secretory fusion partner DsbA, Biotechnology 13 (1995) 982-987.

Conner, G.E., Cathepsin D, in: A.J. Barrett, Rawlings, N.D., Woessner, J.F. (Eds.), Handbook of proteolytic enzymes, vol. 1, Elsevier Ltd., Amsterdam, 2004, pp. 43-52.

Danese, P.N., Silhavy, T.J., Targeting and assembly of periplasmic and outer membrane proteins in *Escherichia coli*, Annu. Rev. Genet. 32 (1998) 59-94.

di Guana, C., Lib, P., Riggsa, P.D., Inouyeb, H., Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein, Gene 67 (1988) 21-30.

Eggers, C.T., Wang, S.X., Fletterick, R.J., Craik, C.S., The role of ecotin dimerization in protease inhibition, J Mol Biol 308 (2001) 975-991.

Erickson-Viitanen, S., Manfredi, J., Viitanen, P., Tribe, D.E., Tritch, R., Hutchinson, C.A., Loeb, D.D., Swanstrom, R., Cleavage of HIV-1 gag polyprotein synthesized in vitro: sequential cleavage by the viral protease, AIDS Res. Hum. Retroviruses 5 (1989) 577-591.

Glockshuber, R., Wunderlich, M., Skerra, A., Rudolph, R., Yield when disulfide-bonded proteins are secreted, USP6,333,175 (2001)

Hellebust, H., Bergseth, S., Örning, L., Expression of the second epidermal growth factor-like domain of human factor VII in *Escherichia coli*, J. Biotechnol. 66 (1998) 203-210.

Hewitt, E.W., Treumann, A., Morrice, N., Tatnell, P.J., Kay, J., Watts, C., Natural processing sites for human cathepsin E and cathepsin D in tetanus toxin. Implications for T cell epitope generation, J. Immunol. 159 (1997) 4693-4699.

Kang, Y., Yoon, J.-W., Effect of modification of connecting peptide of proinsulin on its export, J. Biotechnol. 36 (1994) 45-54.

Lin, X., Universal procedure for refolding of recombinant proteins, USP 6,585,268 (2003)

Liu, G., Topping, T.B., Randall, L.L., Physiological role during export fro the retardation of folding by the leader peptide of maltose-binding protein, Proc. Natl. Acad. Sci. USA 86 (1989) 9213-9217.

Mackin, R.R., Proinsulin: recent observations and controversies, Cell. Mol. Life Sci. 54 (1998) 696-702.

Malik, A., Rudolph, R., Söhling, B., Use of enhanced green fluorescent protein EGFP to determine pepsin at high sensitivity, Anal. Biochem. 340 (2005) 252-258.

Maric, M.A., Taylor, M.D., Blum, J.S., Endosomal aspartic proteases are required for invariant-chain processing, Proc. Natl. Acad. Sci. USA 91 (1994) 2171-2175.

McGrath, M.E., Gillmor, S.A., Fletterick, R.J., Ecotin: Lessons on survival in a protease-fillled world, Protein Sci. 4 (1995) 141-148.

Mergulhao, F.J.M., Taipa, M.A., Cabral, J.M.S., Monteiro, G.A., Evaluation of bottlenecks in proinsulin secretion by *Escherichia coli*, J. Biotechnol. 109 (2004).

Moks, T., Abrahamsen, L., Holmgren, E., Bilich, M., Olsson, A., Uhlen, M., Pohl, G., Sterky, C., Hultberg, H., Josephson, S., Holmgren, A., Jörnvall, H., Nilsson, B., Expression of human insulin-like growth factor I in bacteria: use of optimized gene fusion vectors to facilitate protein purification, Biochemistry 26 (1987) 5239-5244.

Nilsson, B., Abrahamsen, L., Uhlen, M., Immobilization and purification of enzymes with staphylococcal protein A gene fusion vectors, EMBO J. 4 (1985) 1075-1080.

Pál, G., Sziláyi, L., Gráf, L., Stable monomeric form of an originally dimeric serine proteinase inhibitor, ecotin, was constructed via site directed mutagenesis, FEBS Lett. 385 (1996) 165-170.

Raina, S., Missiakas, D., Making and breaking disulfide bonds, Annu. Rev. Microbiol. 51 (1997) 179-202.

Rathore, A.S., Bilbery, R.E., Steinmeyer, D.E., Optimization of an osmotic shock procedure for isolation of a protein product expressed in *E. coli*, Biotechnol. Prog. 19 (2003) 1541-1546.

Rietsch, A., Beckwith, J., The genetics of disulfide bond formation, Annu. Rev. Genet. 32 (1998) 163-184.

Schäffner, J., Winter, J., Rudolph, R., Schwarz, E., Cosecretion of chaperones and low-molecular-size medium additives increases the yield of recombinant disulfide-bridged proteins, Appl. Environ. Microbiol. 67 (2001) 3994-4000.

Schmoldt, H.-U., Wentzel, A., Becker, S., Kolmar, H., A fusion protein system for the recombinant production of short disulfide bond rich cystine knot peptides using barnase as a purification handle, Prot. Expr. Purif. 39 (2005) 82-89.

Scorilas, A., Diamandis, E.P., Levesque, M.A., Papanastasiou-Diamandi, A., Khosravi, M.J., Giai, M., Ponzone, R., Roagna, R., Sismoni, P., López-Otin, C., Immunoenzymatically determined pepsinogen C concentration in breast tumor cytosols: an independent favorable prognostic factor in node-positive patients, Clinical Cancer Res. 5 (1999) 1778-1785.

Sinha, S., Anderson, J.P., Barbour, R., Basi, G.S., Caccavello, R., Davis, D., Doan, M.T., Dovey, H.F., Frigon, N., Hong, J., Jacobson-Croak, K., Jewett, N., Keim, P., Knops, J., Lieberburg, I., Power, M., Tan, H., Tatsuno, G., Tung, J., Schenk, D., Seubert, P., Suomensaari, M., Wang, S., Walker, D., Zhao, J., McConlogue, L., John, V., Purification and cloning of amyloid precursor protein ß-secretase from human brain, Nature 402 (1999) 537-540.

Stoop, A.A., Craik, C.S., Engineering of a macromolecular scaffold to develop specific protease inhibitors, Nature Biotechnol. 21 (2003) 1063-1068.

Summers, R.G., Harris, C.R., Knowles, J.R., A conservative amino acid substitution, arginine for lysine, abolishes export of a hybrid protein in *Escherichia coli*, J. Biol. Chem. 264 (1989) 20082-20086.

Suzuki, F., Murakami, K., Nakamura, Y., Inagami, T., Renin, in: A.J. Barrett, Rawlings, N.D., Woessner, J.F. (Eds.), Handbook of proteolytic enzymes, vol. 1, Elsevier Ltd., Amsterdam, 2004, pp. 54-61.

Vassar, R., Bennett, B.D., Babu-Khan, S., Kahn, S., Mendiaz, E.A., Denis, P., Teplow, D.B., Ross, S., Amarante, P., Loeloff, R., Lou, Y., Fisher, S., Fuller, J., Edenson, S., Lile, J., Jarosinski, M.A., Biere, A.L., Curran, E., Burgess, T., Louis, J.C., Collins, F., Treanor, J., Rogers, G., Citron, M., Beta-secretase cleavage of Alzheimer´s amyloid precursor protein by the transmembrane aspartic protease BACE, Science 286 (1999) 735-741.

Wang, C.-I., Yang, C., Craik, C.S., Isolation of high affinity inhibitor of urokinase-type plasminogen activator by phage display of ecotin, J. Biol. Chem. 270 (1995) 12250-12256.

Winter, J., Neubauer, P., Glockshuber, R., Rudolph, R., Increased production of human proinsulin in the periplasmic space of *Escherichia coli* fusion to DsbA, J. Biotechnol. 84 (2000) 175-185.

CURRICULUM VITAE

Name	**Ajamaluddin Malik**
Date of Birth	May 14th, 1976
Place of Birth	Siddharth Nagar, India
Marital status	Married
Nationality	Indian

EDUCATION

Pre-University

- 1982-1985 Elementary school, Siddharth Nagar, India
- 1985-1991 High School, Gonda, India
- 1991-1993 Intermediate, Basti, India

University

- 1993-1996 Bachelor of Science in Chemistry, Aligarh Muslim University, India
- 1997-1999 Master of Science in Biotechnology, Aligarh Muslim University, India
- 2002-present Ph.D student, Martin Luther University, Germany

RESEARCH EXPERIENCE

Mar 2002–Present	Ph.D student, Martin Luther University, Germany
Jan 2000 – Feb 2002	Research fellow, Aligarh Muslim University, India
July 1998 – Apr 1999	M.Sc. project, Aligarh Muslim University, India
May – June 1998	Summer research assistant, All India Institute of Medical Sciences, India

ACADEMIC AWARDS

- Junior Research Fellowship from University Grant Commission. India, July 2001
- Junior Research Fellowship from Indian Council of Medical Research, India, July 2001
- Studentship from D.B.T (Govt. of India) during July 1997-April 99 for academic excellence.

PUBLICATIONS

1. Malik, A., Jenzsch, M., Lubbert, A., Rudolph, R., & Söhling, B. (2006) " *Periplasmic production of native human proinsulin as a fusion to E. coli ecotin* " (under revision- Protein Exp. and Puri.)

2. Malik, **A.,** Rudolph, R. & Söhling, B. (2006) "*A novel fusion protein system for the production of native human pepsinogen in the bacterial periplasm*". Protein Expression and Purification, 47: 662–671

3. Malik, **A.,** Rudolph, R. & Söhling, B. (2005) "*Use of enhanced green fluorescent protein to determine pepsin at high sensitivity*", Analytical Biochemistry. 340: 252-258

4. Malik, **A**., Khan, A.U. & Lal, S.K. (2002) "*Chemoprotection profiles of Sodium Thiosulfate on Methyl Methane Sulfonate induced mutagenesis of bacteriophage T4"* Medical Science Monitor, 8(6): 212-220

5. Khan, A.U., **Malik, A.** & Ahmad, M. (2001) "*A unique and self- splicing intron in bacteriophage T4*" Indian Journal of Biochemistry and Biophysics, 38: 289-293.

6. Malik, **A.**, Khan, M.A & Khan, A.U. (2000) "*Prevalence of multiple antibiotic resistance and R-plasmid in E.coli isolates of hospital sewage of Aligarh city in India*" Indian Journal of Clinical Biochemistry, 15(2):27-32.

7. Khan, A.U., **Malik, A.**, Khan, M.A. & Naeem, A., (1999) "*Analysis of multiple drug resistance in the E. coli of sewage and drinking water collected from domestic area of Aligarh city.*" J. Microb. World, 1(2): 29-32.

PATENTS

Malik, **A.**, Söhling, B. & Rudolph, R. (2005) - *"Periplasmic production of proteins with therapeutic value by a fusion to a E.coli ecotin*"*- Deutsche Patent- und Markenamt (Patent pending-10 2005 050 732.8-41)

* Title translated from German to English

POSTERS

19th Faltertage, Halle, Germany

22-24 September 2006

"A new fusion protein system for the production of recombinant proteins in the bacterial periplasm"

5th Biotechnology symposium, Leipzig, Germany

18-19 May 2006

"A new fusion protein system for the production of recombinant proteins in the bacterial periplasm" **3rd position- best poster award**.

The protein engineering summit 2006, Boston, USA

24-28 April 2006

"A new fusion protein system for the production of recombinant proteins in the bacterial periplasm"

18th Faltertage, Wittenberg, Germany

7-9 April 2005

"A novel fusion protein system for the production of human pepsinogen in the periplasm of *E. coli*"

DECLARATION

I, the undersigned, declare that this dissertation is solely my own work and no part of it has been submitted to other Universities or Higher Learning Institutions. In addition, all sources of materials used in this dissertation have been duly acknowledged.

Ajamaluddin Malik

Halle/Saale