Turk J Chem
34 (2010) , 1 – 13.

doi:10.3906/kim-0901-37

Cellular distribution of activity for three enzymes with maltose binding protein as fusion partner and the structural implications

Özde F. ÜTKÜR[1,*], Berna SARIYAR AKBULUT [2] Amable HORTAÇSU[1,†]
[1] *Boğaziçi University, Faculty of Engineering, Department of Chemical Engineering, Bebek, 34342, İstanbul-TURKEY*
e-mail: hortacsa@boun.edu.tr, amable.hortacsu@gmail.com
[2] *Marmara University, Faculty of Engineering, Department of Bioengineering, Göztepe Campus, Kadıköy, 34722, İstanbul-TURKEY*

Received 23.01.2009

The bacterial SEC pathway is commonly used for secretion of heterologous proteins in *E. coli* by fusing them to transported proteins to facilitate downstream processing. While some proteins are translocated very efficiently, some reside in the cytoplasm. In this work, maltose binding protein (MBP) was fused to 3 cytoplamic enzymes from *Thermus thermophilus* (serine protease, 251 residues; glucose isomerase, 381 residues; pullulanase, 718 residues) to study the protein transport from the cytoplasm by quantifying the distribution of activities in different cellular compartments. Pullulanase activity was harvested exclusively in the periplasm; however, glucose isomerase activity was harvested exclusively in the cytoplasm. Considerable serine protease activity was found in the periplasm, but after 10 h of induction activity dropped sharply and no activity was found thereafter in either compartment. This was attributed to the instability of the plasmid probably caused by the proteolytic activity of the protease Computations of hypothetical folding rates and secondary structure contents of the proteins showed that folding rates, in addition to alpha-helix and beta-sheet contents of proteins, could be important determinants for efficient translocation by the SEC pathway. These results may give clues to predict whether a protein would be a suitable fusion tail for periplasmic transport with MBP.

Key Words: MBP fusion protein, periplasmic secretion, protein length, secondary structure, folding rate.

*Present address: Department of Biochemical and Chemical Engineering, Dortmund University of Technology, Building G1, Emil-Figge-Strasse 66, D-4 4227 Dortmund, Germany
†Corresponding author

Introduction

All living organisms, no matter how simple or complex, possess the ability to translocate proteins across biological membranes and into different cellular compartments. For this purpose a range of membrane transport processes exist; one of them, the major pathway used to translocate proteins across the bacterial cytoplasmic membrane, is known as the SEC pathway. Over the past 2 decades the SEC pathway has been studied extensively and is well characterised at both genetic and biochemical levels.[1] The SEC translocon, which is defined as the minimal machinery that can accomplish protein transport across the membrane, is able to export unfolded polypeptide chains through the cytoplasmic membrane.[2] The SEC pathway can be divided into 3 distinct but sequential and interdependent stages: targeting, translocation, and release. In the first stage, polypeptide chains are guided to exit sites in the membrane; in the second stage, the exiting chain crosses the lipid bilayer through the translocase; and, in the final stage, the translocated chain is released and allowed either to acquire its native folded state in the periplasm or to proceed to the outer membrane for integration.[3] Proteins transported by this way are initially synthesised as preproteins containing an amino-terminal signal sequence of 18-26 residues, which have 3 characteristic regions: a positively charged residue at the N-terminus, a highly hydrophobic region, and a polar region containing the signal peptidase cleavage site.[2] These signal peptides are cleaved by signal peptidases following the translocation stage. The SEC pathway is able to transport precursor proteins of various lengths, from the cytoplasmic space of *E. coli* by the involvement of the cytoplasmic transport chaperone SecB.[4,5] The most important feature of SEC pathway proteins is that, in addition to a specific signal peptide, they are slow folding proteins, a feature that enables them to be recognised by the chaperone SecB.[4,6] Upon binding, SecB keeps the candidate protein in a competent state for transport. It has been shown in various studies that the signal peptide is essential for the observed slow folding. Nevertheless, its presence alone is not enough for this observation.[7]

A natural SEC pathway substrate, maltose binding protein (MBP) (a periplasmic protein), is commonly used as a fusion partner to target proteins to bacterial secretory routes for transport of the heterologous proteins to different cellular compartments or even to the culture medium.[8–10] In addition to enhancing transport, MBP offers a means of circumventing the inclusion body formation, which is one of the greatest technical obstacles to the production of biologically active proteins in heterologous systems. It has also been demonstrated that MBP can promote the proper folding of its fusion partners.[11]

Translocation of foreign proteins is advantageous for numerous reasons. Above all, the purification costs are significantly reduced. Additionally, if the transported protein has disulphide bonds, the protein will form more stable structures in the periplasm, since these bonds are generally unprocessed in the cytoplasm. Even though not all fusion proteins constructed can be transported, there are no available methods to predict if a candidate fusion protein may be translocated to the periplasmic space.

Based on this knowledge, we fused 3 enzymes to MBP to obtain mechanistic indicators of properties favourable for translocation in *E. coli* by the SEC pathway. In order to determine the ability of the cells to secrete different MBP fusion proteins, the activities of the enzymes in 2 compartments were quantified at different periods after induction. The 3 cytoplasmic enzymes of this work, namely serine protease, glucose isomerase, and pullulanase from *Thermus thermophilus* HB8, had different lengths and secondary structure contents. The experimental and computational results obtained provided us with clues to speculate that translocation efficiencies were related to folding rates and secondary structure contents rather than polypeptide chain lengths.

Materials and methods

Bacterial strains and plasmids

T. thermophilus strain HB8 (ATCC 27634) was purchased from ATCC. *E. coli* XL1 strain (endA1, hsdR17 (rk-, mk+), supE44, thi-1, λ-, recA1, gyrA96, relA1, Δ(lac), [F' proAB, lacIqZ Δ M15, Tn10(tetr)]) was provided by TÜBİTAK, Marmara Research Centre. Plasmid pMAL-p2 was supplied by the manufacturer New England Biolabs (Beverly, MA, USA). Recombinant plasmid pBGI, carrying the MBP-glucose isomerase construct, was provided by Berna Sarıyar Akbulut,[12] and pTPi61, carrying the pullulanase gene, was provided by Prof. Sakai from the University of Shizuoka.

Construction of fusion proteins

The serine protease gene (753 bp, gi:46199306) was amplified from *T. thermophilus* genomic DNA by PCR using the primers SPf and SPr, and the pullulanase gene (2154 bp, gi:14861067) was amplified from recombinant pTPi61 vector by PCR using the primers PUf and PUr. The primers used are listed in Table 1. The PCR products were cloned into pMAL-p2 between the XmnI and EcoRI restriction enzyme sites, downstream from the *mal*E gene of *E. coli*, which encodes MBP, and the vectors were named pOSP and pOPU, respectively (Figure 1). Expression was under the control of *tac* promoter and the *mal*E translation initiation signals.

Table 1. Primers used in this study.

Primer name	Sequence
SPf	5′ – AAGAGAAGGATTTCTAT GCGCG GCCTCGTG – 3′
SPr	5′ – ATTGAATTCTCATACCGCCCCCACT CT– 3′
PUf	5′ – ATAGGAAGGATTTCTATGCTTCACATCAGCCGAA CG – 3′
PUr	5′– AGTGAATTCTCACCCTGCAGGATGGAC – 3′

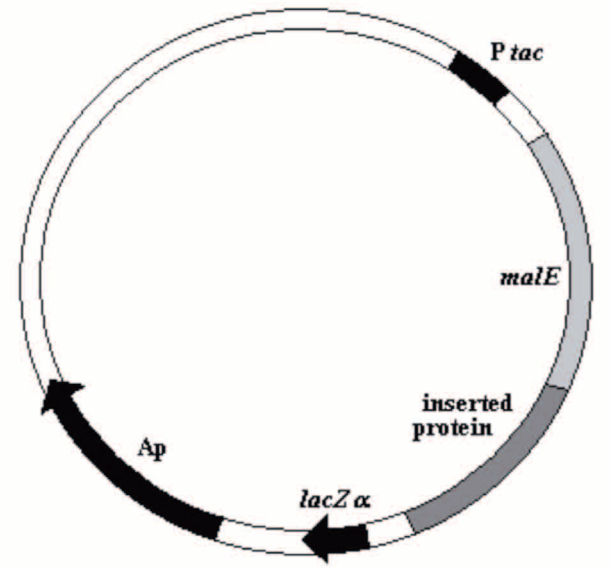

Figure 1. Vector map for the recombinant plasmids.

Cultivation media and conditions

The *E. coli* XL1 cells transformed by pOSP, pOPU, and pBGI were grown in shaking cultures (180 rpm, 37 °C) in LB broth. Culture volume was kept at one-fifth of the flask volume for efficient aeration. Cell growth was monitored spectrophotometrically at OD_{600} nm. The chemical inducer IPTG (isopropyl-β-D-thiogalactopyranoside) was added at a concentration of 1 mM at the mid-exponential phase and the cells were incubated for 20 h after induction. For enzyme analysis 30 mL samples were taken.

Preparation of cell extracts

Samples were harvested by centrifugation at 6000 rpm for 20 min at 4 °C. The periplasmic proteins were released from the precipitated cells by osmotic shock as described by Harrison et al.[13] The harvested cells were resuspended in 0.02 culture volume of 30 mM Tris-HCl containing 20% glucose. Then 1 mM EDTA was added to the suspension, followed by incubation for 20 min at room temperature with gentle stirring. Then the cells were reharvested by centrifugation at 9000 rpm for 20 min at 4 °C and shocked by resuspension and gentle stirring for 20 min in 0.03 culture volume of ice-cold 5 mM $MgSO_4$. Cells, separated from the osmotic shock fluid, were dissolved in 50 mM Tris-HCl buffer. The cytoplasmic cell extract was prepared by ultrasonication of the above cell suspension on ice. Insoluble proteins and membranes were removed from the lysate by centrifugation at 9000 rpm for 30 min at 4 °C. The extracts from the 2 compartments were assayed and the enzyme activity was determined.

Serine protease assay

The assay for the measurement of serine protease activity was adapted from Kazan et al.,[14] in which 500 μL of extract containing serine protease was mixed with 2.5 mL of Tris(hydroxymethyl)-amino methane-maleate (Tris-maleate) buffer (0.2 M of Tris(hydroxymethyl)-amino methane and maleic acid) and 0.2 M NaOH solution supplemented with 0.6% casein solution. The reaction proceeded for 20 min at 30 °C and it was stopped by addition of 2.5 mL of trichloroacetic acid (TCA) solution. Following 30 min incubation at 30 °C, the mixture was filtered and 2.5 mL of Na_2CO_3 was added to the filtrate. The mixture formed by the addition of 0.5 mL of double diluted folin reagent was incubated at room temperature for 30 min. Finally, the change in colour was observed and the absorbance at 660 nm was measured. During the reaction, casein was converted into tyrosine by serine protease in the sample solution, and thus the colour turned to blue. Buffers with pH values between 6.0 and 10.0 were prepared and optimum activity was found in Tris-maleate buffer with a pH of 7.0. This buffer was used for further analysis. One unit of activity was defined as the amount of serine protease that released 1 μg of tyrosine per min under the assay conditions. A calibration curve was prepared with tyrosine as the standard. Then the optical density readings at 660 nm were converted to μg of tyrosine in the reaction mixture. Knowing the amount of tyrosine formed, SP activity was converted to enzyme units.

Glucose isomerase assay

Forty microlitres of extract containing glucose isomerase was mixed with 40 μL of reaction mixture, which contained 0.8 M glucose, 10 μM $MgSO_4$, and 1 μM $CoCl_2$. The reaction proceeded at 65 °C for 30 min and it

was stopped on ice.[15] The amount of fructose formed after the enzyme reaction was estimated by the cysteine-carbazole-sulfuric acid method.[16] Forty microlitres of the reaction tube was added to tubes containing 1.2 mL of 70% H_2SO_4, 40 μL of 0.12% (w/v) carbazole in ethanol, and 40 μL of 1.5% (w/v) cysteine hydrochloride in water. The reaction was allowed to change colour for 10 min at room temperature and the absorbance was read at 560 nm. One unit of activity was defined as the amount of enzyme that released 1 μm of ketose per min under the assay conditions described above.[15]

Pullulanase assay

The dinitrosalicylic acid (DNS) method was used to measure reducing sugar. Five hundred microlitres of extract containing pullulanase was mixed with 500 μL of pullulan solution. The mixture was incubated at 37 °C for 3 min and then 3 mL of DNS solution was added. After incubation in boiling water for 10 min, the mixture was kept on ice and 8 mL of distilled water was added. Finally, the change in colour was observed and the absorbance at 540 nm was measured. During the reaction, pullulan was converted into maltotriose by pullulanase and the colour turned to brown. One unit of activity was defined as the amount of pullulanase that released 1 μg of maltotriose per min under the assay conditions. A calibration chart was prepared with maltotriose as the standard. Then the optical density readings at 540 nm were converted to μg of maltotriose in the reaction mixture. Knowing the amount of malotriose formed, pullulanase activity was converted to enzyme units.

SDS-page analysis

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) of proteins was carried out as described by Laemmli.[17] Following determination of the protein concentration by the method of Bradford,[18] 1.2 μg of protein was loaded onto each lane of the 10% polyacrylamide gels. Gels were stained with colloidal Coomassie Brilliant Blue G250.[19]

Results and discussion

Growth and Expression of the fusion proteins in E. coli XLI

E. coli XLI cells harbouring the recombinant plasmids pOSP, pOPU, and pBGI expressed MBP-serine protease, MBP-pullulanase, and MBP-glucose isomerase fusion proteins, respectively, upon induction with IPTG. The aim was to compare the distribution of activity of 3 cytoplasmic enzymes between the cytoplasm and periplasm using MBP as a fusion partner.

In Figure 2, the growth of *E. coli* XLI cells harbouring the recombinant plasmids was compared with the control culture. There were no significant growth differences between cultures expressing glucose isomerase, pullulanase, and the control culture. However, cell growth was significantly retarded with serine protease expression. Expression without growth continued up to 10 h after induction, after which cell growth resumed. Experiments showed that ampicillin was depleted in the growth media, and plasmid-free cells resumed growth. Further experiments confirmed that this was due to the instability of the plasmid pOSP. The expression of a proteolytic enzyme may be a possible explanation for this cellular behaviour (see below).

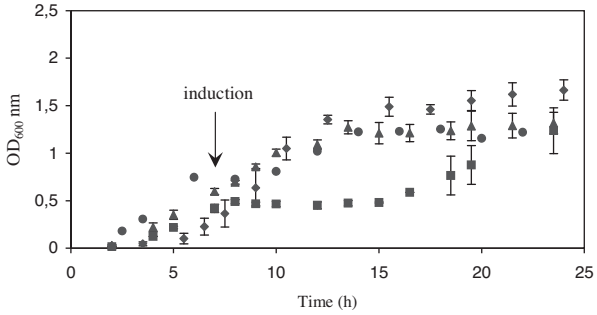

Figure 2. Growth profiles of the *E. coli* XL1 cells harbouring various recombinant plasmids, pMAL-p2 plasmid (control) (▲), pOSP plasmid (■), pBGI plasmid (♦), and pOPU plasmid (●).

When the relative enzyme activities and their distribution over different compartments were analysed, it appeared that the results varied depending on the type of the construct and period of induction. The distribution of activities of the 3 enzymes between the periplasm and cytoplasm after induction are given in Figure 3. Activity of pullulanase, the longest of the 3 proteins studied, shown in Figure 3(C), increased gradually after induction and reached 7756 units/gram dry cell weight (U/gDCW) after 20 h in the periplasm. Pullulanase activity harvested in the cytoplasm was found to be a maximum of 224 U/gDCW after 5 h of induction. Overall about 95% of the total pullulanase activity was harvested in the periplasm. Figure 3(B) shows that most of the active glucose isomerase, the middle sized protein, was retained in the cytoplasm. Maximum glucose isomerase

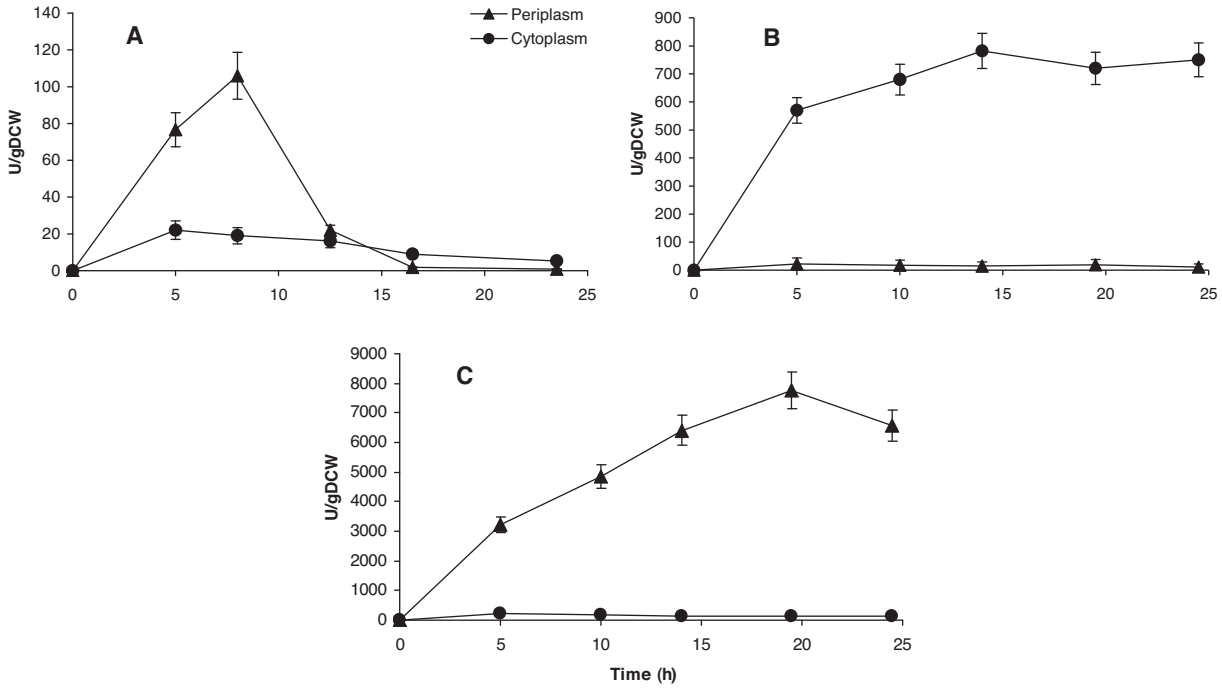

Figure 3. Time dependent activity of MBP fusion proteins after induction in different cellular compartments: (A) serine protease, (B) glucose isomerase, and (C) pullulanase.

activity was 782 U/gDCW in the cytoplasm and 22 U/gDCW in the periplasm. The amount of periplasmic glucose isomerase never exceeded 4% of the total fusion protein amount. Maximum serine protease activity was harvested after 10 h of induction as 106 U/gDCW in the periplasm and which accounted for almost 85% of the total active serine protease. There was an interesting observation with this construct: after 10 h of induction, serine protease activity dropped sharply (Figure 3(A)). It was probably degraded or inactivated due to its proteolytic function. Additionally, since the cells no longer retained the recombinant plasmid pOSP due to its instability after 10 h of induction, there was no further expression of this protein.

Glucose isomerase and pullulanase are enzymes involved in reactions with polysaccharides. When these enzymes have their specific substrates, they convert them or break them down to provide fuel for the cell. However, serine protease is a proteolytic enzyme, the substrates of which are proteins. In its natural host, it may be inactive unless it is required by the cellular functions. However, serine protease expressed in this study could always be active and this continuous proteolytic function may be the cause of growth retardation and hence instability of the plasmid.

Figure 4 shows the percentage distribution of the enzyme activities harvested after 10 h of induction and Figure 5 shows the SDS-PAGE analysis of the periplasmic serine protease and pullulanase. These findings show that most of serine protease and pullulanase activity was localised in the periplasm. Parallel to these findings of activity distributions cytoplasmic pullulanase, cytoplasmic serine protease, and periplasmic glucose isomerase were not detectable on SDS-PAGE.

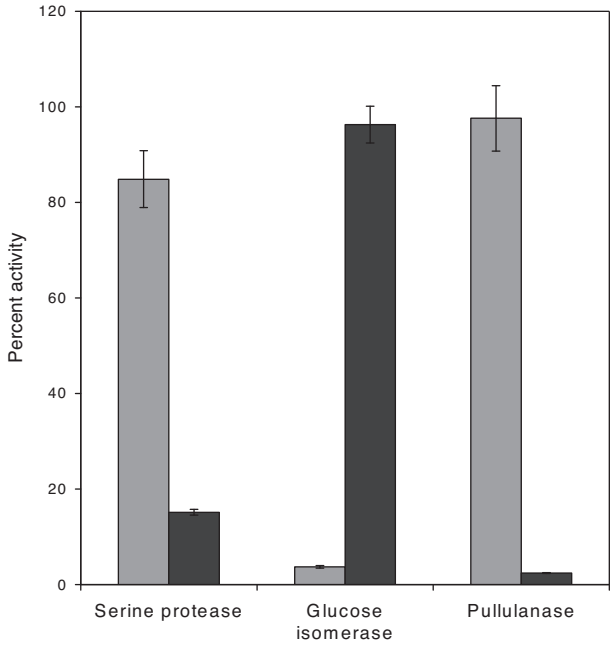

Figure 4. Distribution of activity of MBP fusion proteins between the cytoplasm (■) and periplasm (■) at the end of 10 h of induction.

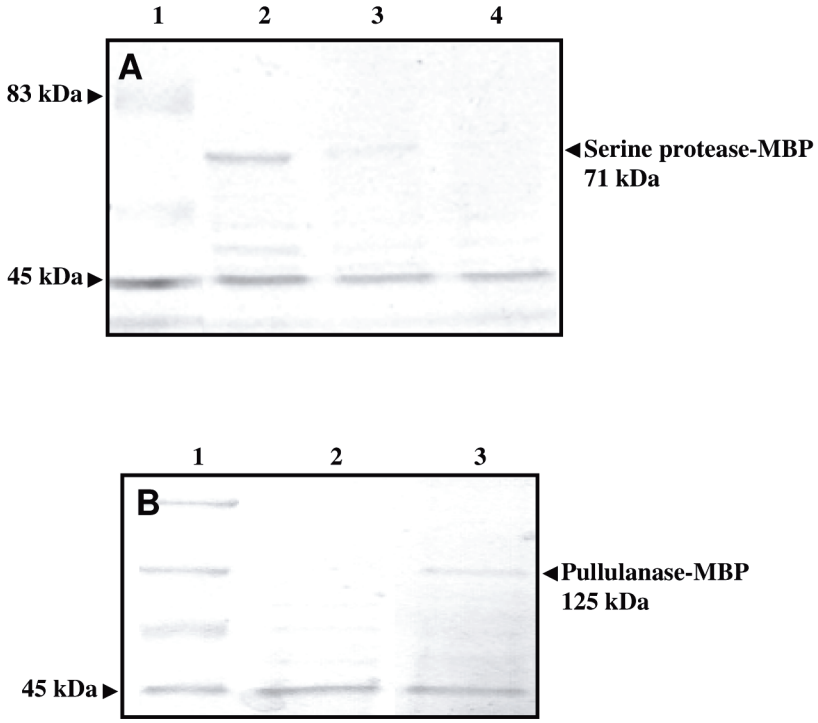

Figure 5. Electrophoretic analysis of the MBP-fusion proteins in the periplasmic space. (A) serine protease: Lane 1: Marker, Lane 2: 10 h after induction, Lane 3: 15 h after induction, Lane 4: 20 h after induction (B) pullulanase: Lane 1: Marker, Lane 2: before induction, Lane 3: 10 h after induction.

Translocation of fusion proteins

Significantly different results were obtained for the 3 cytoplasmic enzymes from the same organism. It was seen that the structural properties of proteins have a stronger influence on translocation through the SEC pathway than the chain length. The activity of the longest and the shortest enzymes, pullulanase and serine protease, respectively, were mainly harvested in the periplasm. On the other hand, the middle sized protein, glucose isomerase, was mainly localised in the cytoplasm. The translocation of this enzyme was negligible, in accordance with previous findings.[12]

These results may partly be explained by specificity of the SEC pathway towards folding rates of the substrate proteins[20] and not on the number of residues. The number of residues may, however, influence the duration of the translocation process once it starts as it is a stepwise process.[21] This effect would have been crucial if the membrane translocation step had been rate limiting. A slow folding rate enhances recognition by the transport chaperone SecB and hence these proteins are directed by SecB to the Sec translocon on the plasma membrane.

The significant correlation between the protein-folding rates and their hierarchical structures suggests that folding rates are largely determined by amino acid sequences.[22–24] Each amino acid is a folding unit and the folding rate of a protein is related to its amino acid sequence, suggesting that the folding rate constant

can be predicted from inter-residue interactions, which are influenced by physical, chemical, energetic, and conformational properties of amino acid residues.[24] Therefore, we wanted to explain the observed experimental results by relating the translocation efficiency to the implications the sequence of residues have on the secondary structure and folding rates of the protein.

Gromiha et al.[24] analysed the relationship between amino acid properties and the folding rates, and developed multiple regression equations for predicting folding rates of different structural classes of proteins.

Their web server at http://psfs.cbrc.jp/fold-rate was used to predict and compare the folding rates of serine protease, glucose isomerase, and pullulanase. In addition to these, 6 SecB substrate proteins, 3 MBP fusion proteins, 1 pelB fusion protein (also transported by SEC pathway), and 4 proteins secreted by pathways other than SEC were analysed. This method required information on the structural family of proteins (alpha, beta, or mixed). Therefore, analysis by Gromiha's web site was mostly restricted to proteins with crystal structures in the Protein Data Bank. For pullulanase and serine protease, the structural families were first estimated using their amino acid sequences (see below for details) and then their folding rates were predicted.

The results in Table 2 show that natural SecB substrates are slow folding proteins. Proteins such as beta-

Table 2. Folding rates predicted for selected proteins.

	Number of residues	PDB ID	Ln(k_f) (s^{-1})
Serine protease	251	-	0.0674
Glucose isomerase	381	1BXB	17.4
Pullulanase	718	-	4.59
MBP [a]	387	1JW4	-0.304
Pectate layase (pelB) [a,*]	375	-	-0.0343
Barstar [a]	90	1A19	-7.01
RNase A [a]	124	1KF2	-3.20
Bovine pancreatic trypsin inhibitor [a]	58	1JV9	-40.2
Levansucrase[a]	473	1OYG	-4.63
Beta-1,4-galactosyltransferase [b]	286	1NMM	-2.93
Pokeweed antiviral protein [b]	261	1GIK	1.48
Variant surface glycoprotein [b,**]	364	1VSG	5.30
Beta-lactamase inhibitory protein [b]	165	2G2U(B)	1.43
Pig alpha-amlyase [c]	496	1PIF	4.55
Beta-lactamase [c]	286	1TML	-3.05
Leader peptidase [c]	234	1BF2	-8,16
Green fluorescent protein [c]	241	1QXT	-2.01

a: Natural SecB substrates,

b: Recombinant proteins studied by fusing to MBP

c: Proteins transported from the cytoplasm by mechanisms other than SEC

Prediction for structural family was based on the sequence of pectate lyase A, C, D, and E.

*As an example variant surface glycoprotein from *Trypanosoma brucei* was analysed.

1,4-galactosyltransferase and pokeweed antiviral protein fused to MBP and beta-lactamase inhibitory protein fused to pelB, which have been successfully transported to the periplasm, are slow folding proteins. Alpha-amylase, beta-lactamase, leader peptidase, and green fluorescent protein, which are transported by different mechanisms, are also slow folding proteins. Therefore, it is not surprising to find out that serine protease and pullulanase are transported to the periplasm. However, glucose isomerase, which has a very fast folding rate compared to the rest of the proteins, could not be transported from the cytoplasm. Likewise, the faster folding variant surface protein from *Tyrpanosoma* was exclusively localised in the cytoplasm when fused to MBP.[25]

The fusion proteins constructed in this study have at least 2 domains[26], 1 of which is MBP. It is possible that SecB recognises its natural substrate MBP and hinders its folding but not that of the protein fused to MBP. Only if the fused protein folds slow enough will it remain competent for translocation until it reaches the membrane. For the glucose isomerase fusion, the results seem to indicate that even if it is recognised by SecB it folds and remains mainly in the cytoplasm before it could reach the translocation machinery. That explains why most of the activity is found in the cytoplasm. The enzyme activity values measured in the cytoplasm in this work are comparable to those found in another study (unpublished results) using pET28a(+), which express cytoplasmic glucose isomerase.

The involvement of a secondary structure in the calculation of folding rates has raised its significance in protein transport. The minimal secondary structure elements, alpha helices and beta sheets[27], may play a fundamental role in the stabilisation of transition state structure components.[28,29] The finding that the folding of the hairpin occurs 30 times slower than the rate of helix formation[27,30–32] supports many experimental and theoretical studies in which it was indicated that alpha proteins have faster folding rates than beta proteins.[24,33–35] Based on this knowledge, we tried to predict the secondary structure contents of serine protease, glucose isomerase, and pullulanase together with those proteins given in Table 2 to make a qualitative correlation for the distribution of activity between the periplasm and cytoplasm.

It is straight forward to get the secondary structure contents of proteins from the Protein Data Bank. However, if the crystal structure of a protein is not available helix and sheet contents may be estimated from the amino acid sequence (correlation is about 80%).[36,37] For this purpose the following web-based servers were used for estimating the secondary structures of the proteins given in Table 2.

1. Protein Structure Prediction Server (PSIPRED) (http://bioinf.cs.ucl.ac.uk/psipred/psiform.html)

2. PredictProtein (PROFsec) (http://www.aber.ac.uk/∼phiwww/prof/)

The predictions were compared with the values from the Protein Data Bank, when available, to show the reliability of the results. Table 3 shows that the results from the 2 methods are comparable for the selected proteins, which ranged from natural SecB substrates to various periplasmic and excreted proteins.

Using structures deposited in the Protein Data Bank, natural SecB substrates such as MBP, bovine pancreatic trypsin inhibitor, and barstar[5], levansucrase, pectate lyase B and RNase A were found to have at least 17% beta-strands. The values calculated by the prediction servers were considerably close to the real values except for some deviation for the helix content of levansucrase. Pokeweed antiviral protein[9] and β-1,4-galactosyltransferase,[10] which were successfully translocated from the cytoplasm when fused to MBP, were also calculated to have at least 19% beta-strand contents using structures in the Protein Data Bank. On the other

Table 3. Structural contents of selected proteins from the Protein Data Bank (when available), and predicted from the 2 web-based servers.

Protein	Method	$\alpha-$helix (%)	$\beta-$sheet (%)
MBP[a]	PROFsec	39	16
	PSIPRED	45	16
	PDB	**43**	**18**
Serine protease	PROFsec	46	14
	PSIPRED	47	18
Glucose isomerase	PROFsec	42	***12***
	PSIPRED	47	***11***
	PDB	**50**	***9***
Pullulanase	PROFsec	15	23
	PSIPRED	21	24
Barstar [a]	PROFsec	40	16
	PSIPRED	57	16
	PDB	**43**	**17**
Bovine pancreatic trypsin inhibitor [a]	PROFsec	12	21
	PSIPRED	17	24
	PDB	**13**	**25**
Levansucrase [a]	PROFsec	7	28
	PSIPRED	7	31
	PDB	**16**	**40**
Pectate lyase B [a,*]	PROFsec	8	33
	PSIPRED	10	33
RNase A [a]	PROFsec	21	26
	PSIPRED	22	23
	PDB	**20**	**35**
Variant surface glycoprotein [b,**]	PROFsec	50	4
	PSIPRED	53	3
	PDB	**50**	**9**
Beta-lactamase inhibitory protein [b]	PROFsec	10	31
	PSIPRED	21	19
	PDB	**20**	**36**
β-1,4-Galactosyl transferase [b]	PROFsec	21	18
	PSIPRED	27	18
	PDB	**25**	**21**
Pokeweed antiviral protein[b]	PROFsec	29	24
	PSIPRED	41	16
	PDB	**37**	**21**

a: Natural SecB substrates, b: Recombinant proteins studied by fusing to MBP
*No PDB structure available, **N-terminal domain

hand, surface variant glycoprotein, which was not translocated when fused to MBP, had only 9% beta sheets (Table 3).

These findings enable us to interpret the experimental results obtained with pullulanase, serine protease, and glucose isomerase fusion proteins. Pullulanase has a high strand content: 23% by PROFsec and 24% by PSIPRED. This explained its high translocation efficiency. The beta-strand content of serine protease is 14% by PROFsec but 18% by PSIPRED, which is comparable to the naturally transported proteins. On the other hand, the beta-strand content of glucose isomerase is well below that of the transported proteins. Glucose isomerase, like surface variant glycoprotein, which was not translocated in *E. coli*, has only 9% beta sheets and 50% alpha helices.

Conclusions

The focus in this study was to determine the ability of cells to secrete different MBP fusion proteins. The results obtained varied considerably between enzymes whose structural motifs seem to correlate well with efficient or inefficient protein transport by the SEC pathway. When SecB binds an unfolded protein, stable folding is prevented and the protein becomes competent for translocation. Therefore, it should be possible to transport cytoplasmic slow-folding proteins when fused to SecB substrates, such as MBP. Glucose isomerase could not be transported from the cytoplasm due to its fast-folding rate. Serine protease and pullulanase, which have folding rates within the range for secreted proteins, were transported.

Folding rates were also correlated with secondary structure contents. Generally, as the number of alpha helices increases, proteins tend to fold faster. Among the 3 proteins, only glucose isomerase has a significantly high helix and low sheet percentage.

There are no available methods to know if a candidate fusion protein may be translocated to the periplasmic space prior to experimentation. The preliminary findings from our work and those obtained from the literature may lead to a secondary structure criterion to predict if a candidate fusion protein may be translocated to the periplasmic space.

Acknowledgements

This work was supported by Boğaziçi University Research Fund Project 03HA501 and Marmara University Research Fund Project FEN-BGS-100105-0063.

References

1. Stephenson, K. *Mol. Membr. Biol.* **2005**, *22*, 17-28.
2. Agarraberes, F. A.; Dice, J. F. *Biochim. Biophys. Acta.* **2001**, *1513*, 1-24.
3. Economou, A. *Mol. Microbiol.* **1998**, *27*, 511-8.
4. Driessen, A. J. M. *Trends. Microbiol.* **2001**, *9*, 193-6.

5. Varadarajan, R.; Panse, V. G. Interaction of Chaperone SecB with Protein Substrates: A Biophysical Study. Indian Institute of Science, Molecular Biophysics Unit, (17-Oct-2006), http://etd.ncsi.iisc.ernet.in/handle/2005/242
6. Driessen, A. J. M.; Fekkes, P.; van der Wolk, J. P. W. *Curr. Opin. Microbiol.* **1998**, *1*, 216-222.
7. Randall, L. L.; Topping, T. B.; Hardy, S. J. S.; Pavlov, M. Y.; Freistroffer, D. V.; Ehrenberg, M. *Proc. Nat. Acad. Sci. USA* **1997**, *94*, 802-7.
8. Ford, C. F.; Suominen, I.; Glatz, C. E. *Protein Expression and Purif.* **1991**, *2*, 95-107.
9. Honjo, E.; Watanabe, K. *Biosci. Biotechnol. Biochem.* **1999**, *63*, 1291-4.
10. Shibatani, S.; Fujiyama, K.; Nishiguchi, S.; Seki, T.; Maekawa, Y. *J. Biosci. Bioeng.* **2001**, *91*, 85-7.
11. Kapust, R. B.; Waugh, D. S. *Protein Science* **1999**, *8*, 1668-1674.
12. Sariyar, B.; Özkan, P.; Kırdar, B.; Hortaçsu, A. *Enzyme. Microb. Technol.* **2004**, *35*, 105-112.
13. Harrison, J. S.; Keshavarz-Moore, E.; Dunhill, P.; Berry, M. J.; Fellinger, A.; Frenken, L. *Biotechnol. Bioeng.* **1996**, *53*, 611-22.
14. Kazan, D.; Denizci A. A.; Kerimak, M. N., Erarslan, A. *J. Ind. Microbiolgy Biotechnology* **2005**, *32*, 335-344.
15. Lee, C.; Bhatnagar, L.; Saha, B. C.; Lee, Y.-E.; Takagi, M.; Imanaka, T.; Bagdasarian, M.; Zeikus, J. G. *Appl. Environ. Microbiol.* **1990**, *56*, 2638-2643.
16. Dishe, Z.; Borenfreund, E. *J. Biol. Chem.* **1951**, *192*, 583-7.
17. Laemmli, U.K. *Nature* **1970**, *227,* 680-5.
18. Bradford, M. M. *Anal. Biochem.* **1976**, *72*, 248-254.
19. Neuhoff, V.; Stamm, R.; Eibl, H. *Electrophoresis* **1985,** *9* 427-448.
20. Randall, L. L.; Hardy, S. J. S. *Trends Biochem. Sci.* **1995**, *20*, 65-9.
21. Schriebel, E.; Driessen; A. J. M.; Hartl, F.-U.; Wickner, W. *Cell* **1991**, *64*, 927-939.
22. Huang, J. T.; Tian, J. *Proteins* **2006**, *63*, 551-4.
23. Ma, B. G.; Guo, J. X.; Zhang, H. Y. *Proteins* **2006,** *65*, 362-72.
24. Gromiha, M. M.; Thangakani, A. M.; Selvaraj, S. *Nucleic Acids Res.* **2006**, *34*, W70-4.
25. Nurcahyo, R. W. Isolierung rekombinanter Varianzglykoproteine von Trypanosoma congolense aus Escherichia coli. [Isolation of recombinant variant surface glycoproteins of *T. congolense* from *E. coli*.] (Summary in English.) Thesis, Freie Universität Berlin, Berlin, Germany, 1998, 113 pp.
26. Smyth, D. R.; Mrozkiewicz, M. K.; Mcgrath, W. J.; Listwan, P.; Kobe, B. *Protein. Sci.* **2003**, *12*, 1313-22.
27. Galzitskaya, O. V.; Higo, J.; Finkelstein, A. V. *Curr. Prot. Pep. Sci.* **2002**, *3*, 191-200.
28. Nölting, B.; Andret, K. *Proteins* **2000**, *41*, 288-98.
29. Guerois, R.; Serrano, L. *Curr. Opin. Struct. Biol.* **2001**, *11*, 101-6.
30. Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196-9.
31. Dyson, H. J.; Rance, M.; Houghten RA, Wright PE, Lerner RA. *J. Mol. Biol.* **1988**, *201*, 201-17.
32. Blanco, F. J.; Rivas, G.; Serrano, L. *Nat. Struct. Biol.* **1994**, *1*, 584-90.
33. Galzitskaya, O. V.; Garbuzynskiy, S. O.; Ivankov, D. N.; Finkelstein, A. V. *Proteins* **2003**, *51*, 162-6.
34. Ivankov, D. N.; Finkelstein, A. V. *Proc. Nat. Acad. Sci. USA* **2004**, *101*, 8942-4.
35. Gromiha, M. M. *J. Chem. Inf. Mod.* **2005**, *45*, 494-501.
36. Dubchak, I.; Muchnik, I.; Holbrook, S. R.; Kim, S. H. *Proc. Nat. Acad. Sci. USA* **1995**, *92*, 8700-4.
37. Bahar, I.; Atilgan, A. R.; Jernigan, R. L.; Erman, B. *Proteins* **1997**, *29*, 172-85.