Identification of factors impeding the production of a single-chain antibody fragment in *Escherichia coli* by comparing *in vivo* and *in vitro* expression

PETER OELSCHLAEGER [1], STEFAN LANGE [1], JUTTA SCHMITT [1], MARTIN SIEMANN-HERZBERG [2,*], MATTHIAS REUSS [2], ROLF D. SCHMID [1,**]

[1] *Institute of Technical Biochemistry, University of Stuttgart, Allmandring 31, 70569 Stuttgart, Germany*

[2] *Institute of Biochemical Engineering, University of Stuttgart, Allmandring 31, 70569 Stuttgart, Germany*

** Communicating author: Prof. Dr. Rolf. D. Schmid

Phone: +49-711-6853193

Fax: +49-711-6853196

E-mail address: Rolf.D.Schmid@rus.uni-stuttgart.de

* for correspondence with regard to cell-free protein synthesis:

Dr. Martin Siemann-Herzberg

Phone: +49-711-6855161

Fax: +49-711-6855164

E-mail address: siemann@ibvt.uni-stuttgart.de

Abstract

In order to produce the atrazine-specific scFv K411B, it was expressed in either the cytoplasm or the periplasm of *Escherichia coli* BL21(DE3). For periplasmic production, the scFv was N-terminally fused to the pelB leader, whereas the unfused variant resulted in cytoplasmic expression. The extent of protein accumulation differed significantly: The expression level of the scFv with leader was 2.3 times higher than that of the protein without leader. To further investigate this, the respective translation profiles were generated by coupled *in vitro* transcription/translation assays and gave according results. Periplasmic expression resulted in only 10% correctly folded scFv. The same percentage was obtained when the scFv was expressed *in vitro*, indicating that the oxidizing environment of the periplasm did not increase proper folding. Thus, the data obtained *in vitro* confirmed the findings observed *in vivo* and suggested that the discrepancy in expression levels was due to different translation efficiencies. However, the *in vivo* production of the scFv with EGFP fused C-terminally (scFv-EGFP) was only successful in the cytoplasm, although *in vitro* the expression with and without the leader rendered the same production profile. This indicated that neither the translation efficiency nor the solubility but other factors impeded periplasmic expression of the fusion protein.

Introduction

Many factors involved in protein folding, translation and transcription may limit the efficient expression of heterologous proteins in *Escherichia coli*. This is particularly true for eucaryotic proteins, such as e.g. single-chain variable antibody fragments (scFv), which require posttranslational modification to exert their proper action. The pros and cons of periplasmic or cytoplasmic expression of antibody fragments in *E. coli* have been intensively discussed (Martineau et al., 1998; Wulfing and Pluckthun, 1994).

On the one hand, secretion of scFv into the periplasm seems to be a very promising strategy, because the oxidizing environment allows the efficient formation of disulfide bonds. Disulfide bonds do not form in the cytosol (Bessette et al., 1999). Therefore cytoplasmic expression of scFvs frequently leads to the formation of inclusion bodies due to the aggregation of malfolded protein. Refolding of these inclusion bodies can be achieved (Sanchez et al., 1999). On the other hand, toxic effects and formation of periplasmic aggregates may counteract the high-level expression of active protein in the periplasm (Wulfing and Pluckthun, 1994).

However, high-level cytoplasmic expression of active scFvs is also possible. Tavladoraki *et al*. were able to show that a scFv folded correctly in the reducing environment of the cytoplasm even without the formation of disulfide bridges (Tavladoraki et al., 1999). In other cases, scFvs have been engineered by random mutagenesis to fold correctly in the cytoplasm (Martineau et al., 1998). Therefore, it can be concluded that it depends very much on the intrinsic properties of the protein that will determine which of

the above-mentioned expression strategies are most suitable for a certain scFv.

Besette *et al.* have improved folding of proteins with disulfide bonds by altering the natural set of thioredoxins and co-expression of the protein disulfide isomerase DsbC in the cytoplasm and found that originally secreted proteins can be folded efficiently in the cytoplasm (Bessette et al., 1999). Improvements in folding of heterologous proteins in the presence of chaperones and/or folding catalysts have been achieved in cell-free translation (Ryabova et al., 1997) as well as in *E. coli* (Georgiou and Valax, 1996; Wall and Pluckthun, 1995). However, some of these proteins need to be expressed in stoichiometric amounts and co-expression of additional proteins complicates production with regard to plasmid stability and induction and therefore the potential to produce the target protein is decreased.

Studies comparing *in vivo* and *in vitro* expression of scFvs have shown that certain scFvs can be expressed well *in vitro* but only in small quantities *in vivo* (Merk et al., 1999). By comparing the expression levels and ratios of active and inactive recombinant protein in both systems, it should be possible to identify some factors which limit the functional expression of scFvs *in vivo*. Such drawbacks are due to either gene expression in terms of transcription and translation, the physico-chemical properties of the environment or global physiological effects. Some factors (transcription efficiency, translation efficiency and spontaneous protein folding) occur both *in vitro* and *in vivo*. Other factors like translocational events or toxic effects only appear *in vivo.*

For such a comparative approach, we have used different constructs of the atrazine-specific scFv K411B (Kramer and Hock, 1996). This scFv contains two disulfide bonds and can be applied for the detection of the herbicide atrazine in ground and drinking water. We compared the expression of the scFv with an N-terminal pelB leader sequence and the respective scFv without leader sequence in *E. coli* and found different expression levels. The influence of the C-terminal fusion of EGFP to the scFv was also investigated. This approach enabled us to identify factors that limit the expression of heterologous protein.

Materials and methods

Microorganisms and growth conditions

Escherichia coli DH5α (Hanahan, 1983) was used for cloning and plasmid propagation, *E. coli* BL21(DE3) (Novagen, Madison, USA) for protein expression. *E. coli* JM109 was used for the preparation of plasmids destined for *in vitro* expression. Cells were cultivated at 37°C or 30°C in LB or synthetic medium (Wilms et al., 2001), each supplemented with 100 mg l^{-1} (m v^{-1}) carbenicillin. For the expression of recombinant protein, shake flask cultures were inoculated 1:100 with overnight cultures, induced with a final concentration of 1 mM isopropyl *β*-D-thiogalactopyranoside (IPTG) at an OD_{600} of 1.0 to 1.2 and harvested after 4 hours of incubation at 30°C.

Plasmids and primers

The plasmid pCANTAB 5E carrying the gene for the scFv K411B was used (Kramer and Hock, 1996). pET20b(+) was purchased from Novagen (Madison, USA). pEGFP containing the sequence coding for EGFP was purchased from Clontech (Palo Alto, USA). Primers were supplied by ARK (Darmstadt, Germany) or MWG-Biotech (Ebersberg, Germany). The following three plasmids were used as control plasmids in the *in vitro* experiments: pIVEX2.1-gfp (carrying a gene coding for EGFP) was purchased from Roche (Mannheim, Germany). pTST101 and pBW8 (carrying genes coding for a MalE-EGFP fusion protein and a homotetrameric hydantoinase, respectively) were kindly provided by Dr. Josef Altenbuchner, Institute of Industrial Genetics, University of Stuttgart, Germany. Plasmids used for *in vitro* studies were purified using the QIAGEN Plasmid Maxi-Kit (QIAGEN, Hilden, Germany).

Cloning procedures

The gene coding for the scFv K411B in pCANTAB 5E was amplified by PCR using the primers 5´-ACTGATATCTCCATGGTCCAGGTGAAACTGCAGCAG-3´ and 5´-GCAGTCGACGAATTCCCGTTTGATTTCCAGCTT-3´, which introduced the restriction sites *Nco*I at the 5´ and *Sal*I at the 3´ end of the gene. PCR amplifications were performed in a Robocycler gradient 40

(Stratagene, Amsterdam, The Netherlands). The purified 777 bp PCR product was cut with *Nco*I and *Sal*I, repurified and ligated into the likewise cut pET20b(+), resulting in the plasmid pETK(+) containing the gene encoding the scFv K411B with a pelB leader at the N-terminus and a His_6-tag at the C-terminus. To obtain pETK(-), in which the sequence coding for the pelB leader sequence was deleted to allow cytoplasmic expression, pETK(+) was digested with *Nde*I and *Nco*I, followed by Klenow fill in and blunt end religation of the 4369 bp fragment. To generate pETKE(+) and pETKE(-), pETK(+) and pETK(-) were digested with *Sal*I, pEGFP with *Nco*I. After Klenow fill in, the linearized plasmids were digested with *Not*I; the resulting 724 bp pEGFP fragment was ligated with the linearized plasmids pETK(+) and pETK(-), respectively. The resulting plasmids, pETKE(+) and pETKE(-), contained the gene coding for the fusion protein consisting of the scFv gene followed directly in frame by the EGFP gene, with and without the pelB leader sequence at the 5´-terminus. To restore the original number of nucleotides between the Shine Dalgarno sequence and the start codon, two additional nucleotides, which were inserted upstream of the start codon in pETK(-) and pETKE(-), were removed with the QuikChange PCR kit (Stratagene, Amsterdam, Netherlands), resulting in pETK(-)-AC and pETKE(-)-AC. PCR was performed according to the manufacturers' instructions using the forward primer 5´-GAAGGAGATATACATATGGTCCAGGTGAAAC-3´ and the corresponding reverse primer with the complementary sequence, followed by *Dpn*I digestion, dialysis against water and transformation. In analogy, two control plasmids, pETK(+)+AC and pETKE(+)+AC, containing two extra nucleotides were created with the forward primer 5´-

GAAGGAGATATACATACATGAAATACCTGCTG-3´ and the corresponding reverse primer. Plasmids were transformed into *E. coli* DH5α or BL21(DE3) by a heat shock method (Chung et al., 1989) or by electroporation (Sambrook et al., 1989) (Gene Pulser and Pulse Controller, Bio-Rad, München, Germany). Transformed cells were grown overnight at 37°C on LB agar plates containing carbenicillin. The plasmids pETK(+), pETK(-), pETKE(+) and pETKE(-) and the translation initiation regions of all used plasmids are shown in **Figure 1.**

DNA sequencing

DNA was sequenced using the fluorescence-based dideoxy DNA cycle method, primers between 18 and 24 nucleotides in length, a Terminator Ready Reaction Mix with Ampli-Taq-Fs (Perkin Elmer, Wellesly, USA) and a 377 DNA sequencing system (Applied Biosystems, Weiterstadt, Germany).

Coupled *in vitro* transcription/translation

Coupled cell-free protein biosynthesis was performed using an S30 bacterial cell extract system generated from *E. coli* A19 according to Pratt (Pratt, 1984) with minor modifications as previously described (Schindler et al., 2000). Batch-wise cell-free transcription/translation was performed at 30°C as reported (Schindler et al., 2000; Schindler et al., 1999). The reaction mixture contained the respective plasmid at a final concentration of 5.6 nM, 2 kU ml^{-1} T7-RNA polymerase, 48 mg ml^{-1} (m v^{-1}) *E. coli*-tRNA, 100 mM

Hepes/KOH, pH 7.6, 2 mM ATP, 1.6 mM GTP, 1 mM CTP, 1 mM UTP, 250 µM of all 20 amino acids, 18.8 µM folinic acid, 1 mg l^{-1} (m v^{-1}) rifampicin, 100 mM KOAc, 18 mM $Mg(OAc)_2$, 1 mM EDTA, 2 mM dithiothreitol, 0.03% (m v^{-1}) sodium azide, and *E. coli* S30 extract at a final protein concentration of 5.9 g l^{-1} (m v^{-1}) (equals to 1.5 µM total ribosome concentration). Forty mM acetyl phosphate and endogenous acetate kinase were used as an energy regeneration system.

Quantification of protein synthesized in vitro

In vitro synthesized protein was estimated from the incorporation of radiolabeled ^{14}C-leucine: 66.7 µM of ^{14}C-leucine (11.7 GBq $mmol^{-1}$, Amersham Pharmacia Biotech, UK) was added to the standard mixture. At respective times, 4 µL aliquots were withdrawn and the concentration of the protein determined by liquid scintillation counting as described previously (Arnold, 2001). Aliquots of the reaction mixture were further analyzed by SDS-PAGE followed by autoradiography according to (Katanaev et al., 1996). In order to separate soluble and insoluble protein fractions, the entire reaction mixture was centrifuged at 10,000 g for 20 min prior to analysis as described in (Katanaev et al., 1996). The autoradiographs were scanned and further analyzed using the ImageMaster VDS-software, package 2.0. Quantification was performed against external calibration standards of 0-100 µM of radiolabeled ^{14}C-Leu.

Cell disruption and sample preparation

E. coli BL21(DE3) cells from shake flask cultures or the bioreactor were harvested by centrifugation (3,000 g, 4°C, 10 min). Total, periplasmic and cytoplasmic extracts were prepared as suggested by the supplier (Novagen). Cytoplasmic extracts were obtained from protoplasts after periplasmic extraction. The insoluble fraction contained the resuspended pellet after cytoplasmic extraction and centrifugation. The resulting cell fractions were either used directly for SDS-PAGE or cell debris was removed by centrifugation to separate insoluble from soluble protein. The resulting samples were analyzed using SDS-PAGE, ELISA, FLISA and the BCA Protein Assay Kit (Pierce, Rockford, USA) or the Bio-Rad Protein Assay Kit (Bio-Rad, München, Germany) following the instructions of the manufacturers. For these assays normalized amounts were applied, always corresponding to the same volume of the original cultivation broth.

SDS-PAGE, blotting and N-terminal sequencing of proteins

Protein samples were separated by SDS-PAGE in 12.5% (m v^{-1}) polyacrylamide gels (Laemmli, 1970) and stained with Coomassie Brilliant Blue R-250 or electro-blotted onto a PVDF membrane (Sequi-Blot™ PVDF Membrane, Trans-Blot® SD, Semi-dry transfer cell, Bio-Rad, Hercules, USA) as described (Coull et al., 1991; Matsudaira, 1987). The protein of interest was excised from the stained membrane and sequenced N-terminally using a gas phase sequencer (Procise™ 491 Protein Sequencer,

785 Programmable Absorbance Detector, 140 C Microgradient System, Applied Biosystems, Weiterstadt, Germany).

Enzyme-linked immunosorbent assay (ELISA) and fluorophor-linked immunosorbent assay (FLISA)

The amount of active scFv in different samples was monitored by non-competitive hapten-immobilized ELISA as described recently (Oelschlaeger et al., 2002). Since at higher concentrations of scFv the ELISA signal showed saturation, the slope in the linear range was taken as an approximation for the amount of functional scFv.

The amount of bifunctional scFv-EGFP fusion protein was determined by hapten-immobilized non-competitive FLISA as reported (Oelschlaeger et al., 2002). In this assay, the EGFP portion of the fusion protein is directly used as indicator of the amount of fusion protein bound. No saturation was observed in this assay, so FLISA signals could be used directly for quantification of functional fusion protein.

Microscopy

Light microscopy and fluorescence microscopy were performed using a Leica DMIRBE microscope (Leica Microsystems, Wetzlar, Germany) and an HCX PL Apo 63x/1.32-0.6 oil immersion objective. Fluorescence setup: excitation: 488 nm with Polychrome 2 (TILL Photonics, München, Germany); dichroic mirror: F53-009 (AHF Analysentechnik, Tübingen, Germany), emission: bandpass filter 535/50 nm (AHF Analysentechnik). Images were taken with a Hamamatsu Orca C4742-95 camera (Hamamatsu

Photonics K.K., Hamamatsu, Japan) and processed with Openlab 2 (Improvision, Coventry, England).

Induced cells were harvested by centrifugation after 2 hours at 30°C, washed in 80 mM PBS buffer and resuspended (5-times concentrated) in 2% (m v^{-1}) paraformaldehyde in 80 mM PBS. After a 20-minute incubation period at 37°C, cells were centrifuged and resuspended (10 times concentrated) in embedding medium and used for microscopy.

Results

Strategy for the detection of factors that limit the heterologous expression of the scFv K411B and the respective scFv-EGFP fusion protein in *E. coli*

The genes coding for two proteins, the atrazine-specific antibody fragment scFv K411B and the fusion protein of the respective scFv and a C-terminal EGFP were cloned into pET20b(+) for periplasmic and cytoplasmic expression in *E. coli* BL21(DE3) resulting in the following vector plasmids:

1) pETK(+), i.e. the scFv K411B gene was inserted in frame between an N-terminal pelB leader sequence and a C-terminal His_6-tag of the vector plasmid pET20b(+),
2) pETK(-), i.e. the coding sequence of the pelB leader was deleted from pETK(+),
3) pETKE(+), i.e. the EGFP gene was fused to the C-terminus of the scFv K411B gene of pETK(+) and
4) pETKE(-), i.e. the EGFP gene was fused to the C-terminus of the scFv K411B gene of pETK(-).

The coding regions of the plasmids are shown in **Figure 1**.

In order to test whether the insertion of two extra nucleotides upstream of the start codon (inserted as a consequence of the cloning procedure) had an effect on translation efficiency, we also constructed the plasmids pETK(-)-AC and pETKE(-)-AC, in which the original sequence

between the Shine Dalgarno sequence and the start codon was restored. As a negative control for the vectors coding for the precursors, two extra nucleotides were inserted into pETK(+) and pETKE(+) upstream of the start codon, resulting in the plasmids pETK(+)+AC and pETKE(+)+AC.

***In vivo* expression of scFv with and without pelB leader and scFv-EGFP fusion protein without pelB leader**

E. coli BL21(DE3) cells were transformed with the plasmids 1) to 4). The strain transformed with pETKE(+) could not be cultivated. The cells transformed with the other three plasmids, respectively, were induced, harvested and cell extracts were prepared to investigate the total expression levels and the accumulation of functional protein in different cell compartments. The phenotypic appearance of induced cells from all 3 strains was determined by light microscopy. In addition, fluorescence microscopy was applied for the strain expressing the scFv-EGFP fusion protein without leader sequence.

The relative amount of heterologous scFv was estimated by SDS-PAGE of total cell lysates **(Fig. 2)**. For pETK(+) (lane 2), a clear band of 29 kDa, which corresponds to the calculated size of the mature scFv, was obtained after expression of scFv with pelB leader sequence. For pETK(-) (lane 4), a band of identical size was observed resulting from the expression of the scFv without pelB and suggesting that the pelB leader sequence of the scFv precursor had been cut off and translocation had occured (lane 2). This was confirmed by N-terminal sequencing. The densitometric evaluation of the bands in lane 2 and lane 4 indicated a 2.3 times greater extent of product

accumulation for the scFv expressed in the periplasm compared to scFv expressed in the cytoplasm.

Equivalent volumes of the culture supernatant (corresponding to the extracellular fraction), the soluble fraction of the periplasmic extract, the soluble fraction of the cytoplasmic extract, prepared from protoplasts separated by centrifugation after periplasmic extraction, and the resuspended pellet, which contained insoluble protein of both cell compartments and the culture supernatant, were applied to ELISA in order to determine the localization of active scFv **(Table 1)** and the protein contents were measure. When summing up the activities in all fractions, the total activity in the strain expressing scFv with leader sequence (pETK(+)) was 1.6 times higher than in the strain expressing scFv without leader sequence (pETK(-)), although the added up protein contents were 1.3 times higher for the latter. The activity in the different fractions differed significantly. For pETK(+), clear signals were obtained in the extracellular fraction (9.6%), the periplasmic fraction (2.9%), the cytoplasmic fraction (42.8%) and the pellet (44.7%). The percentages of protein in these fractions were 8.7%, 16.6%, 64.3% and 10.4%, respectively. For expression of scFv without leader (pETK(-)), no activity was detectable in the medium and almost none in the periplasmic fraction (1.0%). The activity was higher in the cytoplasm (92.5%) and much lower in the pellet (6.5%) compared to the strain expressing scFv with pelB. The corresponding percentages of protein were 2.9% (medium), 4.1% (periplasm), 75.9% (cytoplasm) and 17.1% (pellet). The high protein content in the pellet was mainly due to an uncharacterized protein of ~32 kDa.

Results obtained with cells expressing the scFv-EGFP fusion protein or the scFv without pelB leader were similar to those with cells expressing the scFv without leader. Binding activity of the scFv-domain, measured by FLISA, was only found in the soluble fraction of the cytoplasm (69.0%) and the pellet (31.0%). These fractions contributed 44% and 33% to the total cell protein. Fluorescence could also be detected in the medium and the periplasmic fraction, but as no binding activity could be observed, the signals were certainly not emitted by the bifunctional fusion protein.

It has to be kept in mind that only a small portion of the protein in the pellet was active, probably due to spontaneous refolding upon resuspension, and the total amount of protein was certainly much higher than suggested by the measured activities. We assume that soluble scFv is correctly folded and active. SDS-PAGE and activity measurement of soluble and insoluble fractions of cells expressing scFv with pelB leader in a fed-batch cultivation (data not shown), indicated that the ratio soluble/insoluble scFv was ~1:10 and the activity was comparable in both fractions. Thus, the ratio amount of active scFv/total amount of scFv can be estimated as 1:1 in soluble fractions and as 1:9 (~10%) in insoluble fractions.

In light micoscopy **(Fig. 3A-D)**, the phenotypes of BL21(DE3) cells expressing the scFv with pelB leader (pETK(+)) and BL21(DE3) cells transformed with the empty vector pET20b(+) were comparable. Clear periplasmic inclusion bodies were not observed indicating that insoluble processed scFv was evenly distributed in the periplasm. For cells expressing the scFv and the scFv-EGFP fusion protein without leader (pETK(-) and pETKE(-)), however, knobs were visible at

one or either end of the cylindrically-shaped cells. These knobs probably consisted of cytoplasmic inclusion bodies of malfolded recombinant protein. In fluorescence microscopy of cells expressing mature scFv-EGFP (pETKE(-)) **(Fig. 3E)**, intense fluorescence coincided exactly with these knobs. Obviously, this protein was aggregated, probably mainly due to the malfolded scFv domain, but the EGFP domain conferred some fluorescence. GFP expressed alone has been shown to fold correctly in the cytoplasm and to give bright fluorescence all over the cell in fluorescence microscopy (Cha et al., 2000).

For further investigation, all four plasmids were successfully expressed *in vitro* (cell-free coupled transcription/translation) and again the accumulation of protein and the ratios of soluble and insoluble protein were determined and compared to the *in vivo* results.

In vitro expression of scFv and scFv-EGFP fusion protein with and without pelB leader

The *in vitro* expression of pETK(+) with the pelB leader sequence resulted in a final concentration of 19 µg ml^{-1} of protein, which was 2.7 times higher than the amount obtained for the plasmid pETK(-) (7 µg ml^{-1}) lacking the respective leader sequence **(Fig. 4A)**. A similar ratio was obtained for the *in vitro* expression of pETKE(+) and pETKE(-) (24 and 10 µg ml^{-1}, respectively, **Fig. 4B**). The data indicate that the C-terminal fusion of EGFP to scFv does not significantly influence the expression level, whereas the N-terminal fusion of the pelB leader sequence does.

According to Ringquist *et al*. (Ringquist et al., 1992), the number of nucleotides between the Shine Dalgarno sequence and the start codon might be responsible for the observed differences. However, pETKE(-)-AC and pETK(-)-AC showed virtually the same expression levels as pETKE(-) and pETK(-), whereas the insertion of the two respective nucleotides in pETK(+)+AC and pETKE(+)+AC led to a 25% decrease in productivity when compared to pETK(+) and pETKE(+) (data not shown). Thus, the highest protein yields were obtained for pETK(+) and pETKE(+). These were in the same range as those obtained for three control plasmids pTST101, pBW8 and pIVEX2.1-gfp (~25 μg ml^{-1}, (Arnold et al., 2001; Schindler et al., 2000)). These plasmids encode a MalE-EGFP fusion protein, a homotetrameric hydantoinase and EGFP under control of the T7 promoter, respectively.

The soluble and insoluble fractions of the expressed proteins were separated by centrifugation and subsequently applied to SDS-PAGE **(Fig. 5)** with the intensities being monitored by autoradiography. For the plasmids pETK(+), pETK(-), pETKE(+) and pETKE(-) the percentage amounts of soluble protein were 8%, 15%, 12% and 10%, respectively. Given the accuracy of autoradiographic evaluations, the variation found in these figures was not considered significant. These data show that neither the N-terminal fusion of the pelB leader sequence nor the C-terminal fusion of EGFP have an influence on the folding of the resulting proteins *in vitro*. The obtained ratios of soluble protein were much lower than those obtained when expressing the control plasmids pTST101 (94%), pBW8 (60%) and

pIVEX2.1-gfp (96%). The proteins encoded by these do not have intra-domain disulfide bonds.

Discussion

Cytoplasmic versus periplasmic expression

We expressed the scFv K411B gene, coding for an atrazine-specific antibody fragment, a protein with two disulfide bonds that can be applied for the detection of the herbicide in ground and drinking water (Kramer and Hock, 1996), heterologously in *E. coli* with and without the pelB leader sequence (pETK(+) and pETK(-)) and with and without the C-terminal fusion of the EGFP gene (pETKE(-) and pETK(-)). In doing so, we observed that the yields of total and functional protein differed for the respective constructs. We compared the total expression level and the expression level of functional protein for scFv with and without pelB leader sequence. *In vivo*, the total amount of scFv accumulated after expression of the scFv with pelB leader (pETK(+)) was 2.3 times, the total activity 1.6 times higher than after expression of scFv without leader (pETK(-)) when summing up the activities measured in different fractions after cell extraction.

The major portion of the precursor was translocated into the periplasmic space and processed to mature protein (verified by N-terminal sequencing), followed either by partial cell lysis or secretion of the scFv into the culture supernatant. Most of the protein in the periplasm formed insoluble periplasmic protein that was found in the pellet (insoluble fraction) after the cell extraction procedure. However, no periplasmic inclusion bodies were visible in light microscopy. In contrast, Kipriyanov et al. reported the formation of periplasmic inclusion bodies (Kipriyanov et al.,

1995). In our case, the heterologous protein might have formed small aggregates, which are not visible by light microscopy. A relatively small fraction of active scFv was present in the soluble fraction of the cytoplasmic extract, whereas the majority of the protein was found in the pellet. This protein gave a signal in ELISA, which was surprising, because ELISA data represent only active protein. Probably this observation was due to spontaneous refolding.

No translocation into the periplasm occured in cells expressing scFv and scFv-EGFP fusion protein without leader (pETK(-) and pETKE(-)). In contrast to scFv with leader (pETK(+)), the amount of active protein in the soluble fraction of the cytoplasmic extract was significantly higher than in the resuspended pellet. For the scFv and the scFv-EGFP fusion protein without leader (pETK(-) and pETKE(-)), the soluble cytoplasmic fraction constituted 92.5% and 69.0% of the total activity, respectively. As the cytoplasm of *E. coli* does not favor disulfide bond formation, it is likely that the scFv and the scFv-EGFP in the cytoplasmic fraction and the pellet obtained activity by spontaneous refolding and oxidation during sample preparation and during resuspension in lysis buffer, respectively. Using microscopy, inclusion bodies contributing ~5% to the total cell volume were observed in living cells expressing the proteins without leader. This is in good agreement with the percentage for the total recombinant protein determined by SDS-PAGE **(Figure 2**, lane 4) and sustains the possibility of spontaneous refolding after cell disruption. Although periplasmic aggregates were also present during the preparation of the cytoplasmic extract of cells expressing the scFv in the periplasm and the

total expression level of this protein was higher, the relative activity of the soluble cytoplasmic fraction was relatively small (42.8%). The observations suggest that cytoplasmic aggregates can be renatured more easily in lysis buffer than periplasmic aggregates. This might be due to missing disulfide bond formation in cytoplasmic expression and erroneous disulfide bond formation in periplasmic expression. To test this assumption, we remeasured the activity in the fractions from the expression of scFv with and without pelB leader after incubation at 20°C for three days. Whereas the relative activities in the soluble cytoplasmic fraction and in the resuspended pellet remained almost the same for periplasmic expression (44.9% and 46.8% after three days compared to 42.8% and 44.7% right after sample preparation), they had changed significantly for cytoplasmic expression (61.2% and 35.4% compared to 92.5% and 6.5%), possibly due to further spontaneous refolding of the cytoplasmic inclusion bodies. The malfolded scFv from periplasmic aggregates could not be further renatured in lysis buffer, but with good yields following a procedure similar to that reported by Buchner et al. (Buchner et al., 1992), which includes reduction and re-oxidation (data not shown). Once the scFv had obtained its native state, it proved to be very stable and still showed good activity after one year of storage at 4°C.

Analysis of expression by comparison to *in vitro* results

Many factors can influence the *in vivo* expression of recombinant protein. However, some of these effects can only be seen *in vivo*, while others occur *in vivo* and *in vitro*. The former we define as "physiological effects"

(complex posttranslational events such as translocation inside the cell, processing, assisted protein folding, protein degradation and also toxic effects), which only occur in the living cell, but not or to a much lower extent in cell extracts. In contrast, the latter we consider “intrinsic effects” (transcription and translation efficiency and spontaneous protein folding), which depend on intrinsic properties of the nucleotide sequences of the DNA and mRNA and on the amino acid sequence.

In vivo protein folding is so far not completely understood, but certainly depends on intrinsic properties of the protein, which is stipulated by the amino acid sequence, but also on “physiological effects”. These include the redox potential of the environment and the presence of folding catalysts, which influence the formation of disulfide bridges, and chaperones, which may assist the proper folding of certain proteins. In the expression system used here, “intrinsic effects” are reduced to the translation efficiency and the spontaneous protein folding problem, because derivatives of the vector plasmid pET20b(+) are high-copy plasmids and the T7-promoter expression system is so efficient that mRNA is synthesized much faster than it can be translated by the typical set of ribosomes in *E. coli*. However, this overproduction of mRNA may lead to secondary structures, which are of minor importance in natural procaryotic expression systems in which transcription and translation are directly coupled.

In the following discussion, we attempt to assign the experimentally observed data to either “physiological” or “intrinsic effects”. As stated earlier, data obtained *in vitro* correspond to “intrinsic effects” and data obtained *in vivo* are the result of “intrinsic” and “physiological effects”. We investigated the relative expression level and the solubilities of heterologous

protein. Since the measurements, in particular in the *in vivo* system, are not very precise, we prefer to classify the observed data as high (>66%), moderate (66-16%), or low (<16%) with 100% indicating the optimal observed protein accumulation (*in vivo* ~10% of total cell protein, *in vitro* 24 μg ml^{-1} protein) and 100% of soluble/functional protein. This classification might seem arbitrary. However, values within one class are at least twice as high as the values that were put in the next lower class.

Accumulation of heterologous protein

Table 2 summarizes the *in vitro* and *in vivo* results for the extent of accumulated scFv and scFv-EGFP fusion protein with and without pelB leader sequence, respectively. The excellent concordance of the *in vitro* and *in vivo* data for the scFv with and without leader and scFv-EGFP without leader suggest that the different extents of accumulation observed *in vivo* reflect "intrinsic effects". Due to the efficient promoter and the fact that the amino acid sequence influences the structure, but not the expression level of a protein, here they can be assigned specifically to the nucleotide sequence of the mRNA and translation efficiency. By comparing the amounts of accumulated recombinant protein for pETK(+), and pETKE(+) with pETK(-) and pETK(-)-AC as well as with pETKE(-) and pETKE(-)-AC, we have shown that the two extra nucleotides directly upstream of the start codon had no influence on translation efficiency; this is in contrast to observations made by Ringquist *et al*. (Ringquist et al., 1992).

A plausible explanation for the low expression levels of the genes without the leader sequence might be the presence of the codon GUC at the

+2 position (the codon directly downstream of the start codon). In the constructs containing the pelB leader sequence, the +2 codon is AAA, which is the most common codon at this position in *E. coli* and leads to higher gene expression levels than most other codons (Gren, 1984; Stenstrom et al., 2001a). The codon usage in this region seems to be important for the initiation of translation (Stenstrom et al., 2001b). Furthermore, the secondary structure of the mRNA of the constructs without the pelB leader sequence might have an impact on the recognition of the mRNA by the ribosome and therefore suppress the negative effect of the two additional nucleotides. This is under further investigation and will be published elsewhere.

We noticed severe problems with the *in vivo* expression of the scFv-EGFP fusion protein with leader sequence. As it could be expressed at high levels *in vitro* (24 µg ml^{-1}), this problem can be attributed clearly to “physiological effects”. Such problems have been discussed elsewhere (Casey et al., 2000; Griep et al., 1999) and also toxic effects of other fusion proteins with EGFP have been described (Feilmeier et al., 2000).

Solubility and functionality

The classifications of the solubilities are shown in **Table 3**. For the *in vivo* experiments they refer to the ratio of active to inactive protein after sample preparation, which might not reflect the situation in the living cell because spontaneous refolding might have occured after cell disruption. However, whether a protein can be refolded spontaneously after isolation from the cell may depend on its processing inside the cell.

For the scFv with leader sequence, the *in vivo* data correspond to the *in vitro* data and only approximately 10% of recombinant protein were soluble. Consequently and according to our classification, the low percentage of soluble protein observed in periplasmic expression reflects an intrinsic effect. Too rapid gene expression or volume limitations can be excluded, because *in vitro* other proteins have been expressed at higher levels and with approximately 100% of soluble protein (Arnold et al., 2001, Schindler et al., 2000). In cell-free coupled transcription/translation, no efficient oxidation and formation of disulfide bonds is possible, which is necessary for most scFvs to fold correctly. Directly after the experiment, the insoluble protein was separated by centrifugation, not allowing any further spontaneous refolding. *In vivo* expression of the scFv into the oxidizing environment of the periplasm did not increase the percentage of soluble protein. Here, oxidization might have occurred for the major portion of scFv, but in an erroneous manner. For cytoplasmic expression, relatively high amounts of active scFv (~50%) and scFv-EGFP (~25%) were observed after sample preparation based on comparison of relative activities in soluble and insoluble fractions **(Table 1)**. Since *in vitro* expression yielded only 10% of soluble protein, it is reasonable to assume that *in vivo* initially also the majority of the recombinant protein was malfolded in the non-oxidizing cytoplasm, which was undermined by microscopic observations. However, in contrast to periplasmic expression, the heterologous protein seemed to be kept in a state, which could be more easily refolded to the native state during sample preparation. This included the use of the detergent Triton X-100, incubation with lysozyme under shaking and sonification. During these

steps a part of the protein might have been refolded and oxidized to its native state. The fact that this procedure did not result in a comparable increase of activity for periplasmic aggregates, although the total amount of scFv was higher, we assign to "physiological effects".

Conclusions

We have shown that the quantitative comparison of *in vitro* and *in vivo* expression of an scFv and an scFv-EGFP fusion protein allows the identification of factors that impede the high-level protein production in *E. coli.* Once identified, these factors can be partially eliminated. Here, translation efficiency which is influenced by the mRNA sequence was a limiting factor. The optimization of the nucleotide sequence with regard to the formation of secondary mRNA structures is under further investigation. The formation of insoluble protein could be assigned to intrinsic properties of the protein, although spontaneous refolding into the native state seemed to be favored after expression in the cytoplasm. However, in order to directly obtain all heterologously expressed scFv in a soluble form, protein engineering or the use of genetically engineered strains as described by Besette *et al.* (Bessette et al., 1999) might be necessary. The failure to express the scFv-EGFP fusion protein fused to the pelB leader *in vivo* was likely due to physiological effects, because it was expressed well *in vitro*.

As many recombinant proteins are expressed very poorly in *E. coli* or in an inactive form, a combination of *in vitro* and *in vivo* expression may help to identify factors impeding the expression. Here, the *in vitro* assay could adequately reproduce the *in vivo* expression levels and therefore

might be a practicable tool to screen for good expression vectors before going into the *in vivo* expression system.

Acknowledgements

We thank Sandra Baumann and Kai Scharenweber, Institute of Biochemical Engineering, University of Stuttgart, for their technical assistance in cell-free protein expression, Dr. Annett Burzlaff, Institute of Cell Biology and Immunology, University of Stuttgart, for technical support with microscopy and Dr. Karl Kramer and Dr. Berthold Hock, Technical University of Muenchen at Weihenstephan for the plasmid pCANTAB 5E. Financial support by the "Forschungsschwerpunkt Biosystemtechnik des Landes Baden-Wuerttemberg" is gratefully acknowledged.

References

Arnold S, Siemann, M., Scharnweber, K., Werner, M., Baumann, S., Reuss, M. (2001) Kinetic modeling and simulation of in vitro translcription by phage T7m RNA polymerase. Biotechnol Bioeng 72: 548-561.

Bessette PH, Aslund F, Beckwith J, and Georgiou G (1999) Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. Proc Natl Acad Sci U S A 96: 13703-13708.

Buchner J, Pastan I, and Brinkmann U (1992) A method for increasing the yield of properly folded recombinant fusion proteins: single-chain immunotoxins from renaturation of bacterial inclusion bodies. Anal Biochem 205: 263-270.

Casey JL, Coley AM, Tilley LM, and Foley M (2000) Green fluorescent antibodies: novel in vitro tools. Protein Eng 13: 445-452.

Cha HJ, Wu CF, Valdes JJ, Rao G, and Bentley WE (2000) Observations of green fluorescent protein as a fusion partner in genetically engineered Escherichia coli: monitoring protein expression and solubility. Biotechnol Bioeng 67: 565-574.

Chung CT, Niemela SL, and Miller RH (1989) One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. Proc Natl Acad Sci U S A 86: 2172-2175.

Coull JM, Pappin DJ, Mark J, Aebersold R, and Koster H (1991) Functionalized membrane supports for covalent protein microsequence analysis. Anal Biochem 194: 110-120.

Feilmeier BJ, Iseminger G, Schroeder D, Webber H, and Phillips GJ (2000) Green fluorescent protein functions as a reporter for protein localization in Escherichia coli. J Bacteriol 182: 4068-4076.

Georgiou G, and Valax P (1996) Expression of correctly folded proteins in Escherichia coli. Curr Opin Biotechnol 7: 190-197.

Goodrow MH, Harrison RO, and Hammock BD (1990) Hapten Synthesis, Antibody Development, and Competitive Inhibition Enzyme Immunoassay for *s*-Triazine Herbicides. J Agric Food Chem 38: 990-996.

Gren EJ (1984) Recognition of messenger RNA during translational initiation in Escherichia coli. Biochimie 66: 1-29.

Griep RA, van Twisk C, van der Wolf JM, and Schots A (1999) Fluobodies: green fluorescent single-chain Fv fusion proteins. J Immunol Methods 230: 121-130.

Hanahan D (1983) Studies on transformation of Escherichia coli with plasmids. J Mol Biol 166: 557-580.

Katanaev VL, Spirin AS, Reuss M, and Siemann M (1996) Formation of bacteriophage MS2 infectious units in a cell-free translation system. FEBS Lett 397: 143-148.

Kipriyanov SM, Dubel S, Breitling F, Kontermann RE, Heymann S, and Little M (1995) Bacterial expression and refolding of single-chain Fv fragments with C- terminal cysteines. Cell Biophys 26: 187-204.

Kramer K, and Hock B (1996) Recombinant Single-chain Antibodies against *s*-Triazines. Food & Agricultural Immunology 8: 97-109.

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680-685.

Martineau P, Jones P, and Winter G (1998) Expression of an antibody fragment at high levels in the bacterial cytoplasm. J Mol Biol 280: 117-127.

Matsudaira P (1987) Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J Biol Chem 262: 10035-10038.

Merk H, Stiege W, Tsumoto K, Kumagai I, and Erdmann VA (1999) Cell-free expression of two single-chain monoclonal antibodies against lysozyme: effect of domain arrangement on the expression. J Biochem (Tokyo) 125: 328-333.

Oelschlaeger P, Srikant-Iyer S, Lange S, Schmitt J, Schmid RD (2002) Fluorophor-linked immunosorbent assay: a time- and cost-saving method for the characterization of antibody fragments using a fusion protein of a single-chain antibody fragment and enhanced green fluorescent protein. Anal Biochem 309:27-34.

Pratt JM. (1984) Coupled Transcription-Translation in Prokaryotic Cell-free Systems. In: Hames BD and Higgins SJ (eds) Transcription and Translation: A practical approach. IRL Press, Oxford, England, pp 179-209.

Ringquist S, Shinedling S, Barrick D, Green L, Binkley J, Stormo GD, and Gold L (1992) Translation initiation in Escherichia coli: sequences within the ribosome-binding site. Mol Microbiol 6: 1219-1229.

Ryabova LA, Desplancq D, Spirin AS, and Pluckthun A (1997) Functional antibody production using cell-free translation: effects of protein disulfide isomerase and chaperones. Nat Biotechnol 15: 79-84.

Sambrook J, Fritsch EF, and Maniatis T. (1989) Molecular Cloning, A Laboratory Manual, 2nd edn. Volume 1, Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Sanchez L, Ayala M, Freyre F, Pedroso I, Bell H, Falcon V, and Gavilondo JV (1999) High cytoplasmic expression in E. coli, purification, and in vitro refolding of a single chain Fv antibody fragment against the hepatitis B surface antigen. J Biotechnol 72: 13-20.

Schindler PT, Baumann S, Reuss M, and Siemann M (2000) In vitro coupled transcription translation: effects of modification in lysate preparation on protein composition and biosynthesis activity. Electrophoresis 21: 2606-2609.

Schindler PT, Macherhammer F, Arnold S, Reuss M, and Siemann M (1999) Investigation of translation dynamics under cell-free protein biosynthesis conditions using high-resolution two-dimensional gel electrophoresis. Electrophoresis 20: 806-182.

Stenstrom CM, Holmgren E, and Isaksson LA (2001a) Cooperative effects by the initiation codon and its flanking regions on translation initiation. Gene 273: 259-265.

Stenstrom CM, Jin H, Major LL, Tate WP, and Isaksson LA (2001b) Codon bias at the 3'-side of the initiation codon is correlated with translation initiation efficiency in Escherichia coli. Gene 263: 273-284.

Tavladoraki P, Girotti A, Donini M, Arias FJ, Mancini C, Morea V, Chiaraluce R, Consalvi V, and Benvenuto E (1999) A single-chain antibody fragment is functionally expressed in the cytoplasm of both Escherichia coli and transgenic plants. Eur J Biochem 262: 617-624.

Wall JG, and Pluckthun A (1995) Effects of overexpressing folding modulators on the in vivo folding of heterologous proteins in Escherichia coli. Curr Opin Biotechnol 6: 507-516.

Wilms B, Hauck A, Reuss M, Syldatk C, Mattes R, Siemann M, and Altenbuchner J (2001) High-cell-density fermentation for production of L-N-carbamoylase using an expression system based on the Escherichia coli rhaBAD promoter. Biotechnol Bioeng 73: 95-103.

Wittmann C, and Hock B (1989) Improved enzyme immunoassay for the analysis of *s*-triazines in water samples. Food and Agricultural Immunology 1: 211-224.

Wulfing C, and Pluckthun A (1994) Protein folding in the periplasm of Escherichia coli. Mol Microbiol 12: 685-692.

Figures legends

Fig. 1 Plasmids. The structures of the four investigated constructs are depicted schematically at the top. PT7: T7 polymerase, *sd:* Shine Dalgarno sequence, pelB: pelB leader sequence, his: poly-histidine tag. The sequences of the ribosome binding sites of the four constructs and control plasmids are shown at the bottom. **tic**: translation initiation codon.

Fig. 2 SDS-PAGE of total cell lysates of *E. coli* BL21(DE3) cells. M: Molecular weight marker, the molecular masses (in kDa) of the proteins are indicated on the left. Lanes 1 and 2: BL21(DE3)/pETK(+), lanes 3 and 4: BL21(DE3)/pETK(-), lanes 5 and 6: BL21(DE3)/pET20b(+). Lanes 1, 3 and 5: uninduced, lanes 2, 4 and 6: induced. Arrows indicate the scFv K411B.

Fig. 3 Light (**A-D**) and fluorescence microscopy (**E**). Images of *E. coli* BL21(DE3) cells expressing scFv K411B constructs. **A** Control, BL21(DE3)/pET20b(+), **B** BL21(DE3)/pETK(+), **C** BL21(DE3)/pETK(-), **D** and **E** BL21(DE3)/pETKE(-). **D** and **E** depict the same cells. Arrows indicate the knobs which are discussed in the text.

Fig. 4 Protein expression kinetics of cell-free coupled transcription/translation.
A pETK(+) and pETK(-), coding for the scFv K411B precursor and the mature scFv K411B.
B pETKE(+) and pETKE(-), coding for the scFv K411B-EGFP precursor and the mature scFv K411B-EGFP fusion proteins. Experiments were done twice, standard deviations are indicated.

Fig. 5 Qualitative analysis of cell-free protein synthesis. Samples were loaded onto SDS-polyacrylamide gels after 60 min of coupled *in vitro* transcription/translation synthesis and autoradiographed. The molecular masses (in kDa) of the proteins are indicated on the left and the right. S (supernatant) and P (pellet) indicate the soluble and insoluble fractions of each batch reaction. pIVEX2.1-gfp, pBW8 and pTST101 are control plasmids.

Table 1 Distribution of active recombinant protein in different cell fractions.

Binding of scFv was determined by hapten-immobilized ELISA. Different sample dilutions were used.The slope of the linear range represents the absolute binding value. The absolute binding of the scFv-EGFP fusion protein was determined by hapten-immobilized FLISA. The applied sample concentrations were adjusted, so that each fraction represents an equivalent volume of the culture broth. In the insoluble fractions (sol), the activity corresponds to the total recombinant protein, in the insoluble fractions (insol) it represents about 10% of the total heterologous protein.

cell fraction	scFv with leader		scFv without leader		scFv-EGFP without leader	
	absolute binding	**relative (%)**	absolute binding	**relative (%)**	absolute binding	**relative (%)**
medium	0.0129	**9.6**	0	**0**	0	**0**
periplasm sol	0.0039	**2.9**	0.0009	**1.0**	0	**0**
cytoplasm sol	0.0573	**42.8**	0.0787	**92.5**	296	**69.0**
insol	0.0599	**44.7**	0.0055	**6.5**	133	**31.0**
sum	0.1340	**100.0**	0.0851	**100.0**	429	**100.0**

Table 2 Expression levels of recombinant proteins in *in vitro* and *in vivo* experiments.

	scFv with leader	scFv without leader	scFv-EGFP with leader	scFv-EGFP without leader
in vitro	high	moderate	high	moderate
in vivo	high	moderate	n.d. [1]	moderate [2]
effect	intrinsic	intrinsic	physiological	intrinsic

[1] not determined.

[2] deduced by microscopical analysis of BL21(DE3)/pETKE(-) and BL21(DE3)/pETK(-).

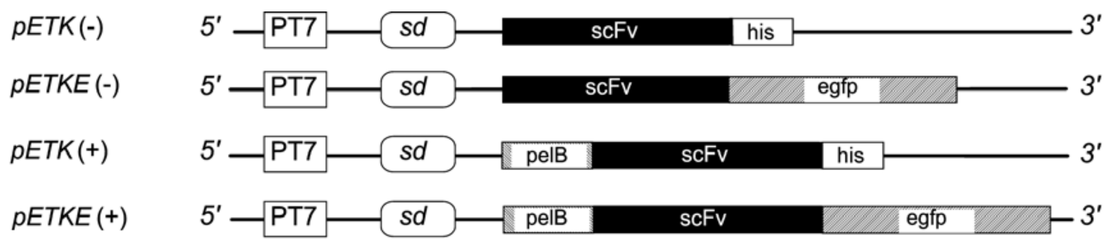
Table 3 Solubility of recombinant proteins in *in vitro* and *in vivo* experiments.

	scFv with leader	scFv without leader	scFv-EGFP with leader	scFv-EGFP without leader
in vitro	low	low	low	low
in vivo	low [2]	moderate [2]	n.d. [1]	moderate [2]
effect	intrinsic	physiological	-	physiological

[1] not determined.

[2] These values were obtained based on the assumption that the content of heterologous protein in the insoluble fraction was about ten times higher than the amount of active protein measured by ELISA and FLISA in the respective samples.

Figure 1:



sd **tic-encoding-gene→**

pETK(-)/pETKE(-)	5´-AGAAGGAGAUAUACAUAC**AUGGUCCAG**-3´
pETK(+)/pETKE(+)	5´-AGAAGGAGAUAUACAU**AUGAAAUAC**-3´
pETK(-)-AC/pETKE(-)-AC	5´-AGAAGGAGAUAUACAU**AUGGUCCAG**-3´
pETK(+)+AC/pETKE(+)+AC	5´-AGAAGGAGAUAUACAUAC**AUGAAAUAC**-3´
pTST101	5´-AGAAGGAGAUAUACAU**AUGAAAACU**-3´
pBW8	5´-AGAAGGAGAUAUACAU**AUGACCCUG**-3´
pIVEX2.1-gfp	5´-AGAAGGAGAUAUACC**AUGACUAGC**-3´

Figure 2:

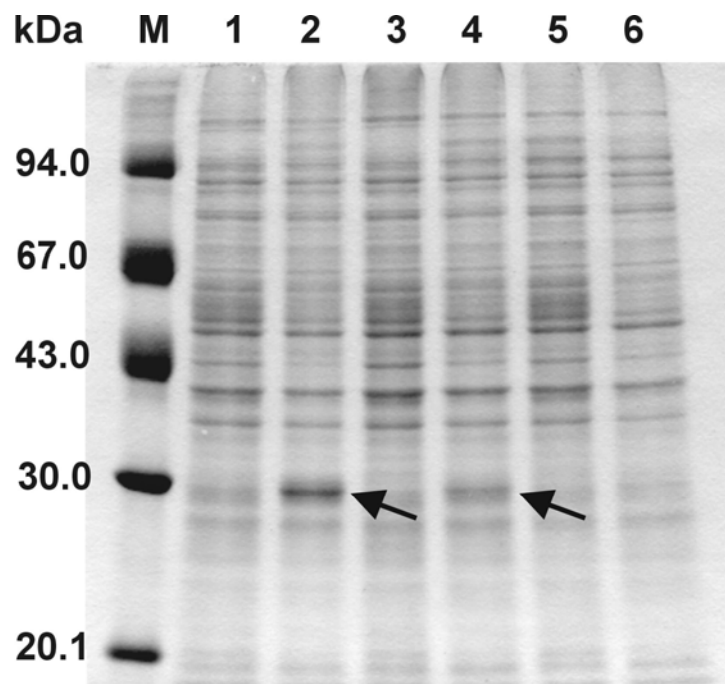


Figure 3:

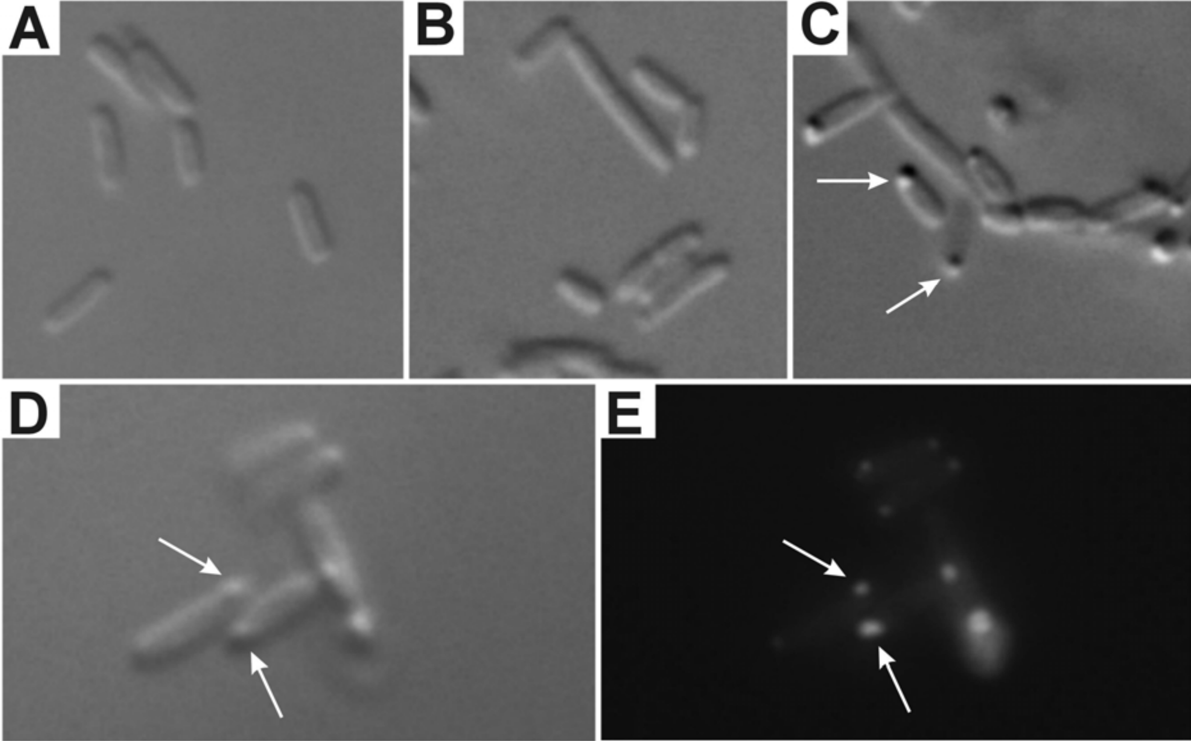


Figure 4:

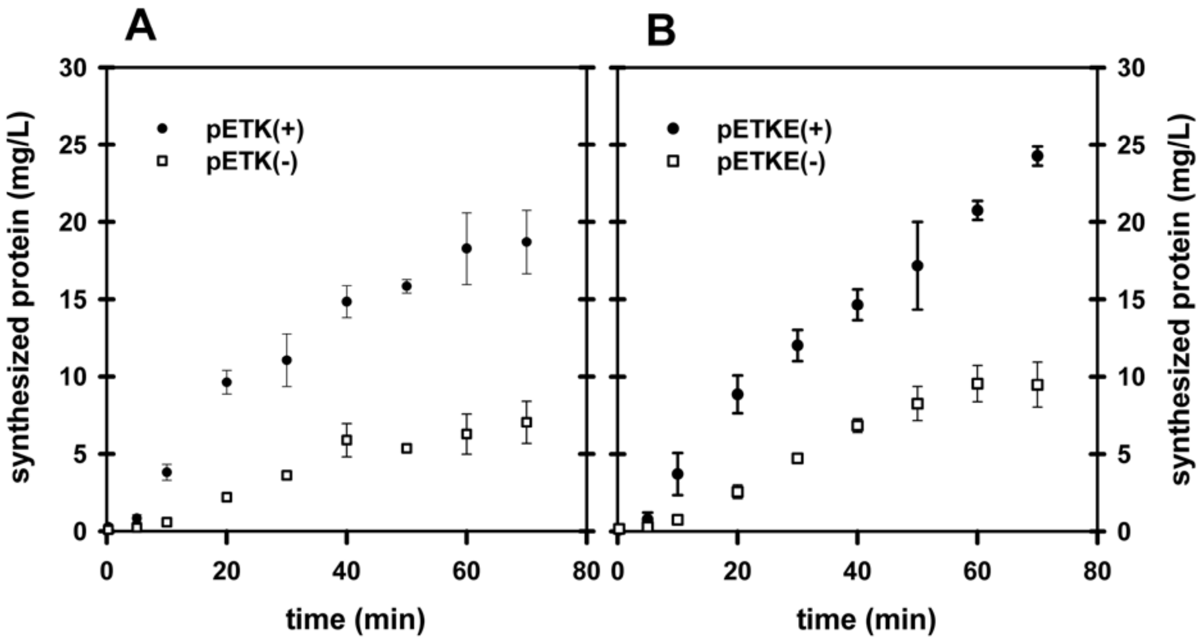


Figure 5: