THESIS ON NATURAL AND EXACT SCIENCES B70

Production of labelled recombinant proteins in fed-batch systems in *Escherichia coli*

KATRIN TOMSON

TALLINN UNIVERSITY OF TECHNOLOGY
Faculty of Chemical and Material Technology
Department of Food Processing

Tallinn University of Technology, Faculty of Science, Department of Gene Technology

National Institute of Chemical Physics and Biophysics, Laboratory of Molecular Genetics, Tallinn, Estonia

Dissertation was accepted for the defence of the degree of Doctor of Philosophy in Natural Sciences on October 5, 2007

Supervisors: Professor Toomas Paalme, Chair of Food Technology, Tallinn University of Technology

Dr. Kalju Vanatalu, Department of Gene Technology, Tallinn University of Technology

Opponents: Dr. Heikki Ojamo, Espoo-Vantaa University of Applied Sciences, Finland

Prof. Jaanus Remme, Institute of Molecular and Cell Biology, University of Tartu, Estonia

Defence of the thesis: December 14, 2007 at Tallinn University of Technology

Members of the jury: Dr. Heikki Ojamo, opponent, Prof. Jaanus Remme, opponent. Prof. Raivo Vilu, Prof. Raivo Vokk, Prof. Priit Kogerman, Prof. Anti Viikna, Dr. Ildar Nisamedtinov

Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology has not been submitted for any academic degree.

ISSN 1406-4723
ISBN 978-9985-59-742-2

3.2.2 B – LOODUS- JA TÄPPISTEADUSED

Märgistatud rekombinantsete valkude tootmine *Escherichia coli* fed-batch süsteemides

KATRIN TOMSON

CONTENTS

INTRODUCTION

The discovery of restriction enzymes (Smith and Wilcox, 1970) that can cut DNA molecules within specific recognition sites made DNA available for manipulation. Following this, development of methods for *in vitro* modification and combination of DNA from different sources (i.e. recombinant DNA methods, genetic engineering) (Jackson et al., 1972; Cohen et al., 1973) started a new era in biology and biotechnology. This allowed for the cloning of genes that code for proteins of interest into the producer cells making it possible to produce proteins that were previously impossible to obtain in quantities required.

Shortly afterwards, many laboratories started the cloning and expression of recombinant genes. Genentech and collaborators were the first to produce human proteins in a bacterium using recombinant DNA technology (somatostatin and insulin in *E. coli*) (Itakura et al., 1977; Goeddel et al., 1979). During the next 30 years, remarkable progress has been made in the field of of recombinant proteins production. In 2006 the list of approved biopharmaceutical products in the United States and Europe numbered 165, including recombinant proteins, monoclonal antibodies and nucleic acid-based drugs (Walsh 2006).

At present recombinant proteins are used in various fields of medicine for diagnosis and therapy, veterinary practice, chemical and food industries and are very widely applied in fundamental research.

Several host cells and organisms of different phylogenetic levels are used to produce recombinant proteins. For a medium or large-scale production, one must carefully choose an expression system and optimise both the production and purification processes. Despite remarkable progress, the cloning of genes and the development of processes to produce a desired protein is far from routine. The entire development process is typically time consuming, consists of many stages with many options, and often relies on trial and error. After all, Nature did not optimise *E. coli* or any other organism to produce proteins for human utilisation (Hoffmann and Rinas, 2004).

One important academic field, where recombinant proteins are increasingly used, is the investigation of protein function and their interactions in complexes with other (macro)molecules. One goal of these studies is promoting the design of new drugs. For studying protein function, the determination of the structure is of vital importance. Generally, the protein under consideration must be native, active, soluble, and of high purity. In addition, proteins labelled with stable isotopes are often required for the elucidation of structure, using, for example, NMR and NS. However, production of the isotopically labelled proteins is one of the most difficult and challenging tasks for heterologous expression in *E. coli* or in any other organism and is regarded as one of the major obstacles of the post-genome era (ESF report, 2001).

This work focuses on the production of (labelled) proteins in the *E. coli* expression system. The following review outlines the factors that influence the

yield and quality of recombinant proteins in practice. All of these factors become even more important when producing labelled proteins as the cost of the feed dramatically increases.

LIST OF PAPERS

This thesis is based on the following papers referred to by their Roman numerals in the text. Some unpublished data are also presented.

I. Tomson K, Liik P and Vilu R (1989) Using of genetically unstable recombinant producers. *Proceedings of the Estonian Academy of Sciences, Biology* 38, 1-8, in Russian

II. Tomson K and Vilu R (1990) Stability of recombinant *Escherichia coli* in continuous culture. *Proceedings of the Estonian Academy of Sciences, Biology* 39, 93-95

III. Tomson K, Paalme T, Laakso PS and Vilu R (1995) Automatic laboratory-scale fed-batch procedure for production of recombinant proteins using inducible expression systems of *Escherichia coli*. *Biotechnology techniques* 9, 793-798

IV. Vanatalu K, Tomson K, Paalme T, Tiisma K and Vilu R (1998) Application of dense culture for *in vivo* deuteration of microorganisms. *Proceedings* of the Workshop on Deuteration of Biological Molecules for Structural and Dynamic Studies. Application to Neutron Scattering and NMR. May 19-25, 1998, Dubna, Russia, 7-12

V. Tomson K, Barber J, Vanatalu K (2006) Adaptastat - a new method for optimising of bacterial growth conditions in continuous culture: Interactive substrate limitation based on dissolved oxygen measurement. *Journal of Microbiological Methods* 64, 380-390, Epub 2005 Jul 1

VI. Patent application: Method of cultivation and recombinant protein expression in *Escherichia coli*; Owner: Tallinn University of Technology; Authors: Kalju Vanatalu, Katrin Tomson; Priority number: US60/ 827, 336; Priority date: 29.09.2006

VII. Massad T, Jarvet J, Tanner R, Tomson K, Smirnova J, Palumaa P, Sugai M, Kohno T, Vanatalu K, Damberg P (2007) Maximum entropy reconstruction of joint phi, psi-distribution with a coil-library prior: the backbone conformation of the peptide hormone motilin in aqueous solution from phi and psi-dependent J-couplings. *Journal of Bimolecular NMR* 38, 107-123, Epub 2007 Apr 26

ACKNOWLEDGEMENTS

This work was carried out at the Laboratory of Molecular Genetics of the National Institute of Chemical Physics and Biophysics.

The financial support for this research provided by Wellcome Trust (Collaborative Research Initiative Grant no. 060781) and by Estonian Science Foundation (grants no. 2310 and 5138) is gratefully acknowledged.

I would like to thank

- My supervisors Prof. Toomas Paalme and Dr. Kalju Vanatalu for teamwork, discussions, support and last but not least, for critical reading of my writing.
- Dr. Kalju Vanatalu, MSc Julia Smirnova, Prof. Peep Palumaa and Dr. Risto Tanner whom I worked together with in the "decathlon" of motilin production.
- Dr. Jüri Jarvet for cooperation and NMR spectra.
- My previous and present colleagues for support and friendliness.

1 LITERATURE REVIEW

1.1 Expression systems

An expression system is a combination of host and vector, which provides a genetic context for expressing a cloned gene to produce the desired protein in the host cell. The hosts used for production of recombinant proteins range from simple prokaryotic bacteria to eukaryotic multicellular transgenic plants and animals, including unicellular eukaryotic organisms (yeast) and the more complex insect and mammalian cells.

1.1.1 The choice of host

The choice of host for production depends on the properties and the final use of the expressed protein. Criteria such as protein structure (complexity, S-S bridges, post-translational modifications), cultivation, down-stream processing, usage, yield/ cost, time, approval, etc to the selection of an expression host are well summarized by Gellissen et al. (2005).

E. coli expression systems were the first to be utilised and are invariably the most used systems for producing recombinant proteins. However, the ability of *E. coli* cells to correctly process and secrete over-expressed recombinant proteins is limited. Therefore, complex proteins needing post-translational modifications for authenticity and/ or biological activity are often produced in alternative expression systems (Dyck et al., 2003; Müller et al., 2006). *E. coli* expression systems are preferred for the production of less-complex recombinant proteins if their folding is not very complex and glucosylation is not required. The main advantages of *E. coli* systems are a variety of available expression vectors and host strains as well as its ability to grow on simple defined culture media. The main disadvantages of *E. coli* systems have been addressed, and the processing ability of *E. coli* has been significantly improved making it an even more attractive host.

For the production of ^{2}H-, ^{13}C- and ^{15}N-compounds, including recombinant proteins, the ability to grow on simple labelled substrates of lower price is extremely important. In addition, an almost complete replacement of H with ^{2}H in water and substrates should be possible. Eukaryotic expression systems are unsuitable for the production of completely deuterated proteins, because eukaryotic cells, with a few exceptions (see below), are not able to grow in fully deuterated medium. Even for production of ^{15}N- and ^{13}C-compounds, the substrate requirements are too complex and product yields too low to obtain the product at a reasonable price using mammalian and insect cell lines.

The ability to grow on several compounds as a sole source of carbon and energy as well as the ability to grow in completely deuterated environments are the big advantages of *E. coli* expression systems for the production of isotopically labelled polypeptides and proteins. Thus, *E. coli* was chosen as the host organism in this study. Reviewing the *E. coli* expression systems, let us briefly cover alternative host/ systems used for labelling.

1.1.2 Alternative hosts/ systems for labelling

A lower eukaryote, the methylotrophic yeast *Pichia pastoris* can grow in fully deuterated medium on methanol (Haon et al., 1993) and has been used for production of recombinant proteins (review by Ilgen et al., 2005). Production of partially deuterated proteins (in a medium with 95% $^{2}H_2O$ and 99% deuterated methanol) was reported by Morgan et al. (2000).

Desplancq et al. (2005) cultivated photoautotrophic cyanobacterium *Anabaena* sp. in a medium consisting of inexpensive ^{15}N-, ^{13}C-labeled $NaHCO_3$ and $NaNO_3$ as the sole sources of carbon and nitrogen and achieved high expression levels of N-terminal domain of *E. coli* gyrase B subunit upon induction of endogenous promoter of the nitrate assimilation with $NaNO_3$. They also observed that *Anabaena* cells can be adapted to grow in medium containing $^{2}H_2O$ without loss of GyrB (1-219) expression, suggesting that it may be possible to overexpress ^{2}H/ ^{15}N/ ^{13}C-labeled proteins in this strain.

In addition, *in vitro* expression systems can be used for synthesis of labelled proteins (Swartz, 2001; Torizawa et al., 2004; Carver and Keniry, 2002). *In vitro* expression systems use a crude *E. coli* fraction as a source of ribosomes and other factors necessary for the synthesis of proteins, and T7 RNA polymerase for transcription. Either dialysis, continuous, or fed-batch technology is used to maintain levels of amino acids and nucleotides and remove the products. Cell-free methods appear to introduce a new degree of complexity (Swartz, 2001), however, in specific cases such as toxic proteins, or the production of many proteins in parallel, or for the selective labelling of proteins, this method could be advantageous. The benefits are, that protein synthesis and folding can be better controlled, expensive labelling substrates can be used only for protein and not for biomass synthesis and the label is not scrambled in the case of selective labelling. In addition, proteins can be synthesised from PCR fragments directly (Nakano et al., 1999).

For the production of recombinant proteins in larger amounts and for uniform labelling, more feasible in vivo methods are generally used.

1.1.3 *E. coli* expression systems

Expression systems of *E. coli* are attractive due to their ability to grow fast and reach high cell densities on defined inexpensive media (Baneyx, 1999). *E. coli* is the model organism for research and a huge base of knowledge and experience has accumulated on its growth and genetics.

The first complete genome sequence was published for *E. coli* reference strain K-12 MG1655 by Blattner et al. (1997). Thereafter, genomes of the popular laboratory strain K-12 W3110 and pathogenic strain O157 were sequenced, the sequencing of the popular host for protein production *E. coli* B being in progress (see NCBI database). All this facilitates further studies, genetic engineering and use of this organism as a production host. *E. coli* has also become a popular expression host due to its growing number of available cloning/ expression vectors and host strains with different properties. However, one major drawback of using *E. coli* as an expression system is misfolding, therefore several expression systems enabling expression of proteins in a soluble form have been developed (Bessette et al., 1999). In case of *E. coli* expression system, the cost of processing and refolding of the proteins can be significant compared to the cost of expression (Hellwig et al., 2004). Processes and costs of production of tissue plasminogen activator (tPA) in *E. coli* and mammalian CHO cell systems were compared and analysed by Datar (Datar et al., 1993) showing the feasibility of *E. coli*-based production of a complex protein such as tPA, once every step is properly optimised.

Often the protein is expressed in a form of inclusion bodies and obtained via a renaturation process; classical examples are the production of Eli Lilly's human insulin (1982, the first recombinant pharmaceutical) and Monsanto's bovine growth hormone (1994) (Kroef et al., 1989; Storrs and Przybycien, 1991).

In any case, the goal of a production process is to obtain a high yield of biologically authentic or active recombinant protein at a reasonable cost and in a reasonable period. For successful production of a recombinant protein in *E. coli* the stability of the expression vector, transcriptional and translational efficiency, stability of the transcribed mRNA, localisation, proteolytic stability and folding of the target protein, as well as cell growth are essential (Jonasson et al., 2002). When using expensive labelled compounds as feedstocks, one must choose the expression system and processing conditions especially carefully with a greater emphasis on optimisation.

1.1.4 Expression vectors

Plasmids are common tools used in cloning, modification and expression of recombinant genes in bacterial systems. Plasmids are small double stranded extra-chromosomal circular DNA molecules capable of replicating independently in a host cell and typically carry gene encoding antibiotic resistance protein. Despite of

the large number of available expression vectors and host strains, systems with new possibilities are continuously being developed.

Essential features and genetic elements of expression vectors

In addition to the gene that codes for the protein of interest, several genetic elements with appropriate sequences must be properly positioned on the expression vector to achieve high level of synthesis (reviewed in Makrides, 1996; Baneyx, 1999; Jonasson et al., 2002): (1) origin of replication (*ori)*, (2) multiple cloning site (MCS) for convenient insertion of the heterologous gene, (3) an expression cassette (containing promoter, transcription terminator, sequences determining secondary structure of mRNA, ribosome binding site (RBS), translation start and stop signals) for efficient and regulated transcription and translation of a gene of interest, and (4) a gene coding for selection marker (e.g. antibiotics) for phenotypic selection and maintenance of the vector in the cell. Plasmids can carry also additional features, like plasmid stability functions and the elements for secretion (secretion signals), better solubility (fusion partners), purification (affinity tags) and detection of the target protein (reporters).

Plasmids have an essential impact on productivity. Related factors are plasmid copy number, structural plasmid stability, and segregational plasmid stability, reviewed in detail by Friehs (2004). DNA segregation is the process of equally distributing replicated genetic material into daughter progeny.

Origin of replication and plasmid copy number

The copy number determined by the origin of replication is an important parameter of the expression system. Most common cloning and expression vectors (e.g. pBR322 and related plasmids, the pUC plasmids, the pET series, the pBluescript series) have the origin of replication derived from ColE1 plasmid or ColE1-type plasmids (pMB1, p15A, and others) naturally occurring in *E. coli* (Actis et al., 1999).

To achieve high gene dosage, heterologous cDNAs are typically cloned into medium or high copy number plasmids that replicate in a relaxed fashion (pMB1/ ColE1-derived plasmids are present at 15-60 and the pUC series of pMB1 derivatives at few hundred copies per cell, respectively). These multicopy plasmids are replicated in a stable manner and randomly distributed during cell division (Baneyx, 1999).

1.1.5 Plasmid stability

The development of plasmid-free cells or mutated plasmids can decrease the yield of recombinant proteins.

Plasmid stability can be divided into structural and segregational stability. Structural stability exists when all generated plasmids have an identical base sequence. The system is segregationally stable when each daughter cell gets at least one copy of the plasmid during cell division. In conditions where plasmids are non-essential to the cells, plasmid-free daughter cells can arise during cell division at a

low frequency that is inversely dependent on plasmid copy number (Summers, 1998). For the same average plasmid copy number, different growth conditions can generate different distributions, dramatically affecting loss rates (Summers, 1991). Efficient replication control reduces variation in copy number and increases segregational stability for a given average copy number (Paulsson and Ehrenberg, 1998).

Modelling has been used to analyse effects of plasmid instability on culture behaviour (Friehs, 2004). In paper I, two models were analysed. The first was a simple model based on two types of cells i.e. cells with and without the plasmid(s). It was shown that in the case of continuous culture in nonselective growth conditions the percentage of plasmid-containing cells decreases exponentially until steady-state culture with only plasmid-free cells present is achieved. Whereas in selective conditions, in steady state the ratio of cells with and without plasmids is determined only by the probability of plasmid loss (p). In the latter case, parameters of recombinant culture such as p, maximal specific growth rate (μ_{max+}), and the saturation constant (K_{s+}) can be calculated from cultivation data. In reality, the situation can be more complicated. Therefore, a segregated model was developed, in which probability of plasmid loss and specific growth rate both depended on plasmid copy number in cells (replication of plasmids once in a cell cycle and random segregation of them between daughter cells was assumed). Computer simulation showed that (1) steady state without plasmid-containing cells is achieved in nonselective conditions whereas (2) stationary distribution of cells as a function of plasmid copy number (determined by values of specific growth rates) is developed in selective conditions.

Metabolic burden, overgrowth

On the other hand, high copy number causes an additional metabolic burden on the plasmid-carrying cells imposed by plasmid replication and target gene expression. This in turn decreases the growth rate allowing the plasmid-free cells to overgrow the culture (Betenbaugh et al., 1989; Bentley et al., 1990; Glick, 1995; Corchero and Villaverde, 1998). If the specific growth rate of plasmid-free cells in current growth conditions is higher than that of plasmid-carrying cells they will overgrow the whole population. At the end of cultivation, there might be a lot of biomass but no plasmid and subsequently no recombinant product. Therefore, different approaches are used to enhance segregational plasmid stability. This is relatively easy to achieve in shake flasks, but is a true challenge in bioreactors or industrial scale cultivations (Friehs, 2004).

Methods for enhancing stability

(1) Applying selective pressure is a widely used method for identifying of transformants and for giving growth advantage to recombinant cells. The most common antibiotic resistance markers in expression plasmids confer resistance to ampicillin, kanamycin, chloramphenicol or tetracycline. Resistance toward penicillins is based on degradation of the β-lactam ring by β-lactamase enzymes,

products of the *bla* genes. However, due to β-lactamase secreted into the medium both ampicillin and carbenicillin are degraded already after few generations of cell growth and selective pressure is lost (Korpimäki et al., 2003). The results shown in (Korpimäki et al., 2003) suggest that ampicillin, carbenicillin (often recommended as less sensitive to low pH/ degradation e.g. Sambrook et al., 1989; Novagen pET System Manual, 11th Edition), or other selection systems based on degradation or inactivation of antibiotics (e.g. chloramphenicol and kanamycin) are not desirable especially in long cultivations. Kanamycin is inactivated in the periplasm by aminoglycoside (phospho-, acetyl-, and nucleotidyl)transferases and chloramphenicol by the *cat* gene product, chloramphenicol acetyl transferase (Ausubel et al., 1987). Tetracycline resistance used in expression plasmids is based on active export of the antibiotic from the cell. Tetracycline was shown to be very stable in the culture medium and therefore being an efficient selection agent (Korpimäki et al., 2003).

Further drawbacks of this approach are the possible impact of the spread of drug resistance to environmental organisms (by accidental gene transfer) and, in particular, pathogens (Williams et al., 1998), and the contamination of the product or biomass by antibiotics, which may be unacceptable from a medical standpoint (Baneyx, 1999). Despite these drawbacks, antibiotics are still widely used for historical reasons, e.g. many plasmid vectors, including many commercially available vectors, carry the gene for β-lactamase *bla* (Friehs, 2004).

(2) Complementation of mutated essential genes. An essential chromosomal gene is deleted or mutated and an intact copy or suppressor is supplied on a plasmid. Plasmid loss leads to cell death under non-permissive growth conditions. Examples include deletion of genes necessary for the synthesis of amino acids, and thermosensitive and nonsense mutations in essential chromosomal genes (Baneyx, 1999). Fiedler and Skerra (2001) used the proline auxotrophy of the common host strain *E. coli* JM83 to stabilize a plasmid containing the *proBA* genes. No plasmid loss was observed during growth and induction phases. The complementation method also has disadvantages: (1) restrictions to the growth medium composition, (2) revertants, where the gene function is restored can arise. These are serious problems for an industrial system.

(3) Titration of the LacI repressor protein. Williams et al. (1998) constructed an *E. coli* host strain containing a conditionally essential chromosomal gene (*kan*) under the control of the *lac* operator/ promoter. In the absence of an inducer (allolactose or IPTG) this strain cannot grow on kanamycin-containing media due to the repression of *kan* expression by LacI protein. Transformation with a high copy-number plasmid containing the *lac* operator (*lac*O) effectively induces *kan* expression by titrating LacI from the chromosomal operator.

(4) Reducing size of plasmid and cloned fragment. In general, natural low-copy-number plasmids are larger than high-copy-number plasmids, which may help to minimise the metabolic burden imposed by the plasmids on their host cells (Ebersbach and Gerdes, 2005). Therefore, to achieve high copy number (gene dosage) the expression plasmids usually contain only genetic elements necessary

for plasmid replication and target gene expression. The plasmid maintenance sequences (see following sections) found on naturally occurring plasmids are commonly not included into the expression vectors.

(5) Post-segregational killing mechanisms (called also programmed cell death or addiction mechanisms) are frequently found on low-copy number plasmids and even on chromosomes (Altenbuchner and Mattes, 2005). Such mechanisms are based on a stable toxin and an unstable antidote. The antidotes neutralize corresponding toxins or prevent their synthesis. In plasmid-free cells the antidotes decay faster than toxins, whereupon toxins are activated and cells are killed. Example of this system, the 580 base pairs long *parB* stability locus of plasmid R1, consists of *hok* (host killing) and *sok* (suppressor of host killing) genes (Gerdes, 1988). Pecota et al. (1997) have found that the use of this antisense killer locus (*hok*+/ *sok*+) in combination with a proteic killer locus (*parDE*+) had an unsubstantial impact on specific growth rate, maintained high β-galactosidase expression and led to a 30 fold increase in the number of stable generations. Effect of plasmids containing *hok*+/ *sok*+ alone was ninefold. This method can introduce a metabolic burden on the cell by requiring transcription, and often translation, of additional plasmid-encoded genes, especially in the case of multicopy-plasmids, which may lead to slow growth. In our studies, no improvement in expression stability was obtained using the hok-sok system (see Stability testing).

(6) Active partition systems. Various active partition systems (*par*-regions) can be found on many natural plasmids. Plasmids and bacterial chromosomes are partitioned in a mitotic-like way in prokaryotes (Møller-Jensen et al., 2000). *Par*-regions can be used to stabilise expression plasmids, as was shown with the runaway plasmid by Friehs and colleagues. They found that stability of the plasmids with and without *par*-region was practically 100% after 100 h continuous cultivation in LB-medium and 0% after 70 h, respectively. The disadvantage of stabilising plasmids with *par*-regions is the significant increase of the plasmid size (Friehs, 2004).

(7) Multimer resolution sequences. ColE1-type plasmids require host polymerases RNAP and DNAP I for replication. Replication is controlled by an anti-sense RNA (RNA I) that binds to the precursor of the primer RNA (RNA II), thus inhibiting Rnase H-mediated primer maturation. A small protein called Rop encoded by the plasmid binds to the complex of RNA I and RNA II and stabilises it (Tomizawa, 1990).

An important cause of multicopy plasmid instability is multimer formation by homologous recombination (in *recA*$^+$ strains, as commonly used expression hosts W3110, BL21, and many others). Replication control by multiple origins of replication on multimers is confused, and multimer-containing cells contain fewer independent plasmids. Furthermore, dimers replicate at twice the rate of monomers and accumulate rapidly in the descendants of the cell in which they first appeared. This generates a sub-population of dimer-only cells from which plasmid-free cells arise at high frequency, the so-called dimer catastrophe (Summers et al., 1993).

In the natural multicopy plasmid ColE1 monomers are regenerated and the normal copy number is restored due to recombination between *cer* sites in a plasmid multimer by the XerCD recombinase (Summers and Sherratt, 1984). For the stable maintenance of ColE1 and related plasmids an additional mechanism is also essential. A promoter within *cer* directs the synthesis of a short transcript (Rcd) in multimer-containing cells. The Rcd checkpoint hypothesis proposes that Rcd delays cell division until multimer resolution is complete (Chant and Summers, 2007).

Most of the cloning and expression plasmids lack the functional *cer* sequence and partition mechanism (e.g. *par*) (Altenbuchner and Mattes, 2005). Observations have shown that strains in which multimer resolution is non-functional are extremely unstable (Summers and Sherratt, 1984; Summers et al., 1993; Boe and Tolker-Nielsen, 1997). Multimer resolution was restored and stability was improved in cells with *cer*-sequence cloned (Wilms et al., 2001).

(8) Changing the plasmid copy number. Positive effects of higher copy number are considered greater stability of the plasmid at random partition at cell division and higher gene dosage, which usually leads to higher expression level. However, especially in the case of strong promoters, toxic or interfering with host metabolism target proteins, use of high copy number plasmids could lead to lower production due to higher instability of plasmid, stress or even cell death. During prolonged high-cell-density cultivation or in continuous processes faster growing nonproducing plasmid-free cells overgrow the culture (Friehs, 2004). Therefore, in such cases expression plasmids with low or inducible copy number may give better results. Obtaining a reasonable expression level with a low copy number could be compensated by stronger promoters. Expression vectors with a dual regulation of both the plasmid copy number and gene expression are described. These vectors are maintained as a single-copy plasmid during growth, and therefore, any residual expression is much more tightly regulated than for the conventional multicopy expression vectors. Commercially available copy-control expression vector pETcoco has alternative origins of DNA replication (oriS for single copy and oriV for medium copy plasmids) that allows regulation of the plasmid amplification (from 1 to 40 copies) and the cloned gene expression based on the pT7lacO promoter (Wild and Szybalski, 2004).

(9) Integration into the chromosome. One possibility to get rid of plasmid instability is the insertion of heterologous genes into the *E. coli* chromosome. Different integration systems are used. Integration via homologous recombination in *rec*-proficient strains (Link et al., 1997), λ-site specific recombination (Atlung et al., 1991), or direct integration of linear PCR-fragment flanked by 40-60 bp long sequences homologous to a chromosomal target in *E. coli* strains that express *recET*. The last method could be used also for engineering of host strain, and for cloning into plasmids without using the restrictases or ligases, called ET-cloning (Zhang et al., 1998).

Effects of cultivation on segregational stability

Lot of data can be found in the literature on this subject (refs in Friehs, 2004). Metabolic burden (or toxicity) and stress caused by expression of recombinant proteins always reduces plasmid stability, and therefore tightly regulatable promoters are used, which allows one to separate cultivation into two phases: growth phase without expression and production phase after induction (Friehs, 2004). The effect of cultivation conditions (DO, pH, t, media/ substrates, feeding methods) on plasmid stability is apparently mediated by complex mechanisms: growth rate and stress.

Amplification of ColE1-type plasmids. Copy number of ColE1-type vectors (e.g. pET vectors) can rise during overproduction of proteins from about 40 to 250 and more (Altenbuchner and Mattes, 2005). Depletion of amino acids leads to the accumulation of uncharged tRNAs. Due to sequence homologies between tRNAs and RNA I (Yavachev and Ivanov, 1988), the interaction between RNA I and RNA II is impaired and replication control is lost (see regulation of replication of ColE1-type plasmids, p 16). The increase in copy number/ gene dosage may eventually exhaust the cell's metabolic capacity and in turn cause breakdown of protein synthesis. In a new ColE1-type vector in which the region of RNA I that most probably interacts with uncharged tRNAs has been altered, the copy number remained constant during high-level protein synthesis (Grabherr et al., 2002).

Structural stability

Various mutations can change the base sequence of the plasmid with various results. If mutations occur in regulatory or structural genes of recombinant protein, the yield and quality of that protein could be severely affected: either less protein is expressed or only protein fragments may be produced. Due to frame shift mutations even proteins with wrong amino acid sequences and perhaps unknown activities, can be synthesised (Friehs, 2004). In normal growth conditions mutation rate is low. In adverse growth conditions, bacteria induce several stress mechanisms: SOS, heat-shock, stringent and the general stress response that completely change their expression pattern to relieve stress. These mechanisms increase the mutation rate, which could give some bacteria in the population a beneficial mutation that enhances survival. These inducible adaptive mutations can be caused by (1) recombination, (2) error-prone DNA polymerase and 3) stress responses that regulate cellular processes. In addition, a subpopulation of cells enters into a state of hypermutation, giving rise to about 10% of the single mutants and virtually all of the mutants with multiple mutations (Foster, 2005).

Structural stability can be increased by avoiding adverse growth conditions (e.g. growing to stationary phase, starvation, intensive expression of recombinant proteins) that induce stress response.

Recently it was shown, that one cause of instability in BL21(DE3) strains is mutation of T7 polymerase gene (Vethanayagam and Flower, 2005).

1.1.6 Promoters and common expression systems

Promoters are DNA sequences, which direct RNA polymerase (RNAP) binding and affect the yield of recombinant protein by determining the rate of transcription. *E. coli* promoters consist of two -10 and -35 hexameric sequences, separated by a 16 - 19 bp spacer.

A number of promoters for high-level protein expression in *E. coli* are reviewed in (Altenbuchner and Mattes, 2005), however only a few of them are commonly used. An effective promoter must be easily inducible and result in the expression level of up to 10 - 30% of total cellular protein, with minimal basal expression level (i.e., it is tightly regulated). Minimisation of basal transcription is especially important when the expression of target introduces a cellular stress and thereby selects for plasmid loss (Sørensen and Mortensen, 2005). In addition, homogeneous (all cells are induced) and graded (expression level is proportional to inducer concentration) induction is a desirable feature of the expression system.

lac, tac, and trc promoters

E. coli lactose utilisation is a well-known regulation mechanism. *Lac* promoter from this operon is rather weak and not commonly used for recombinant protein production. The strong hybrid *tac* and *trc* promoters consist of the −35 region of the *trp* promoter and the −10 region of the *lac*UV5 promoter, and differ only by one bp in the length of spacer domain. All these promoters are repressed by homotetrameric *lac* repressor protein (LacI) and are inducible by non-hydrolysable lactose analog isopropyl-β-D-1-thiolgalactopyranoside (IPTG) that can freely cross the cell membrane. These promoters are also regulated by cAMP receptor protein (CRP) complex (catabolite repression). cAMP is produced as a response to low glucose level (see below).

T7 RNA polymerase system

One of the most widely used expression systems is the T7 RNA polymerase (T7 RNAP) system (Studier and Moffatt 1986). For example, about 80% of the proteins whose three-dimensional structures were submitted to the protein data bank (PDB) in 2003 were produced in an *E. coli* expression system. Whereby T7 based pET expression system represents more than 90% of the 2003 PDB protein preparation systems (Sørensen and Mortensen, 2005). The pET vector has a bacteriophage T7 promoter, which is transcribed only by T7 RNAP and could be used in a host strain carrying a chromosomal T7 RNAP gene under the *lacUV5* promoter, which is 2.5 times stronger and less sensitive to catabolite regulation than *lac* promoter is. LacI repressor protein regulates the *lacUV5* promoter. T7 RNAP gene is transcribed by *E. coli* RNAP when IPTG binds to the repressor LacI and causes its release from the *lac* operator. Subsequently, transcription of the target gene is initiated only by T7 RNAP from T7 promoter. T7 RNAP transcribes up to 230 nucleotides per second and is five times faster than *E. coli* RNAP (Studier and Moffatt, 1986).

Basal expression (leakage) is a common problem in most expression vectors. There are several measures used to deal with this problem in pET vectors.

(1) Using T7*lac* hybrid promoter, repressable by LacI instead of T7 promoter. A copy of the lacI gene is present on the chromosome and on the plasmids with T7*lac* hybrid promoters (Dubendorff and Studier, 1991). (2) Coexpression of T7 lysozyme, a natural inhibitor of T7 RNAP using plasmids pLysS or pLysE (Studier, 1991). This method reduces the background expression about tenfold, but still does not completely remove it. T7 lysozyme is a bifunctional protein that also cleaves the bond between N-acetylmuramic acid and l-alanine in peptidoglycan layer of the *E. coli* cell wall (Inouye et al., 1973), which seems to be reason for lower growth rates and expression yields. (3) Avoiding catabolite activation by adding 0.5 - 1% glucose into the medium or *cya* mutants (Grossman et al., 1998). Such high concentration of glucose can be used only in initial phases for selection of cultures/ colonies but not in high density cultivation, at least without somehow removing the overflow metabolism (see Cultivation). Leaky induction is significantly reduced in cAMP-deficient *cya* mutants that can be used for clone selection and fermentation to avoid counter-selection of plasmids carrying toxic genes under T7-control. (4) The degree of *lac* repression depends upon the ratio of repressing LacI molecules in a cell to the *lac* operator sites on DNA. Enhancement of repression could be achieved when the overexpressing promoter up-mutant strain $LacI^q$ or $LacI^{Q1}$ is employed (Calos, 1978; Glascock and Weickert, 1998). (5) Using expression vectors with inducible copy number (e.g. pETcoco) basal expression during growth phase could be reduced to about 1/ 40 of the levels obtained with pET vectors. This seems to be the best solution to avoid additional burden caused by basal expression of target gene as well as by multi-copy plasmid maintenance, plus constitutive or read-through expression of other genes (as antibiotic resistance, *lac* repressor etc) on the plasmid.

Phage λ promoter pL

Phage λ promoter pL that is regulated by the cI repressor is widely used for protein overexpression. The temperature-sensitive cI857 repressor allows control of gene expression by changing the growth temperature. The cI857 repressor is functional at 30 °C, but inactivated at 42 °C, where gene expression is induced (Love et al., 1996). A constitutive expression system with deleted repressor gene was used for the production of proteolytically susceptible proteins at lower than 42ºC temperatures (25, 30 and 37ºC) (Menart et al., 2003).

Transcription termination

A transcription terminator downstream of the coding sequence enhances plasmid stability by preventing transcription through the replication region and through other promoters located on the plasmid (Hannig and Makrides, 1998).

Messenger RNA, translation and mRNA stability

Translation initiation. Translation is initiated by binding of ribosomes at the ribosomal binding site (RBS) containing the Shine-Dalgarno (SD) sequence, and starts from the initiation codon (Sørensen and Mortensen, 2005). For efficient

translation initiation, several features of mRNA are important (Jonasson et al., 2002): (1) efficient start codon is AUG, (2) optimal SD sequence is UAAGGAGG, (3) SD is located 7 ± 2 nucleotides upstream from start codon, (4) secondary structure of RBS is highly important, (5) high contents of adenine and thymine in this region, (6) codon following the start codon (adenine in highly expressed genes).

Translation elongation and codon usage – the gene. Most amino acids are encoded by more than one codon, and different species have their own preferences in the usage of synonymous codons (codon dialect). The summary of codon usages can be found at http://www.kazusa.or.jp/codon/. The codons AGG, AGA, CGA (Arg); CUA (Leu); AUA (Ile) and CCC (Pro) are very rarely used in *E. coli*, and this is reflected by the level of cognate tRNAs available in the cytoplasm (Dong et al., 1996). Differences in codon usage can impede translation of heterologous proteins due to insufficient tRNAs in the expression host. This can lead to translational stalling, premature translation termination, translational frameshift and amino acid mis-incorporation (Kurland and Gallant, 1996). Two strategies are used with various success to remedy codon bias (Sørensen and Mortensen, 2005; Terpe, 2006): (1) re-design of rare codons in target gene, (2) providing rare tRNAs either by coexpression of rare tRNA genes from codon+ plasmids (e.g. pR.A.R.E, Novagen and CodonPlus plasmids, Stratagene) or using codon+ strains engineered to supply additional tRNAs under control of their native promoters (e.g. commercially available BL21 CodonPlus-RIL, CodonPlus-RP, Rosetta, Rosetta-gami strains).

Similar problems can also be created by the high-level expression of proteins having an abundant amino acid. In this case, supplying the limiting amino acid to the culture medium can improve the expression (Kane, 1995).

Expression of some heterologous genes in *E. coli* can result in the immediate cessation of cell division and in the accumulation of modified or truncated forms of target proteins (Zahn, 1996). An RNA molecule, that acts both as a tRNA and mRNA, designated as tmRNA, directs the modification of proteins of which biosynthesis has stalled or has been interrupted. Translation is then switched from mRNA to tmRNA, and a peptide tag (AANDENYALAA) is added to the truncated protein. The tagged protein is rapidly degraded by Lon protease (Choy et al., 2007). This is a post-translational proofreading system (Altenbuchner and Mattes 2005).

Translation termination. Product heterogeneity can also be caused by incorrect translation termination. Preferred stop codon in *E. coli* is UAA. Frequently, UGA is used, which can lead to translational read-through. Consecutive stop codons or prolonged stop codons UAAU should be used to increase the efficiency of translation termination (Poole et al., 1995).

Stability of mRNA is an important parameter affecting expression level. mRNAs are degraded by endonuclease RnaseE , 3'-5'exonucleases and their complex called degradasome. Commonly used transcription terminators (e.g. TΦ from phage T7) stabilize the mRNA by forming a stem loop at the 3'-end (Newbury et al., 1987;

Jonasson et al., 2002). Similarly, stability could be enhanced by stem-loop formed at the 5'-end, which further cannot be recognized and cleaved by degradasome (Baker and Mackie, 2003). Higher translation rates have resulted in higher mRNA stability. It is assumed that translating ribosomes hide the RnaseE cleavage site and therefore stabilize mRNA (Altenbuchner and Mattes, 2005). RnaseE is essential for the maturation of ribosomal 5S RNA. C-terminal half of RnaseE (needed for formation of degradasome) can be deleted without affecting the maturation function whereas the half-life of (all) mRNAs in the cell is considerably increased by this deletion (*rne* strains, e.g. BL21star, Invitrogen) (Altenbuchner and Mattes, 2005).

1.1.7 Production of proteins

Inner and outer membranes divide *E. coli* cell into two compartments: the cytoplasm and the periplasm. The decision whether to direct recombinant protein into one of these compartments or to the extracellular medium depends on balancing the advantages and disadvantages of each location, and is protein-specific.

Production in cytoplasm

In *E. coli*, expression in the cytoplasm is normally preferred due to higher production yields. Biological activity depends on protein folding that is reviewed in (Baneyx and Mujacic, 2004). In cytoplasm the concentration of macromolecules can reach up to 400 mg ml^{-1} (Ellis and Minton, 2003) and folding under such macromolecular crowding is quite complicated. Misfolding of host proteins can result from premature termination of translation, failure of newly synthesised proteins to reach a native conformation or from loss of structure due to environmental stress.

Cells have developed largely conserved conformational quality control mechanisms to favor *de novo* folding, refold partially folded proteins, dissolve aggregates and degrade irretrievably damaged proteins. Small (<100 residues) single domain host proteins usually fold easily, but many large multidomain and overexpressed recombinant proteins need the assistance of folding helpers as folding catalysts and chaperones. Chaperones are constitutively expressed under balanced growth, but many of them are upregulated upon stresses that increase cellular protein misfolding, and are therefore termed stress or heat shock proteins. Chaperones can be divided into three classes: folding chaperones (e.g. DnaK and GroEL) unfold/ refold proteins in ATP-dependent process, holding chaperones (e.g. IbpB) bind to partially folded proteins, disaggregating chaperone ClpB promotes the solubilisation of aggregated proteins.

In *E. coli* cytoplasm, folding is assisted by three chaperone systems: trigger factor (TF), DnaK-DnaJ-GrpE and GroEL-GroES. TF binds to ribosome near the peptide exit site and protects newly synthesized chains of primarily large (>60 kDa) multidomain proteins, which represent 10 - 20% of the *E. coli* proteome (Deuerling et al., 2003). Longer newly synthesised proteins may alternatively be

bound by DnaK-DnaJ. Both chaperones engage exposed hydrophobic sites that are normally found inside the folded proteins, shielding them from the solvent and each other. After release from TF or GrpE-mediated release from DnaK, protein may reach native conformation, rebind to DnaK-DnaJ or transferred to GroEL-GroES. Under normal conditions, 10-15% of newly synthesized host proteins up to 60 kDa in size are folded in a diluted environment inside the GroEL-GroES chaperone complex. Under heat stress, this part can reach to 30% or more (Ewalt et al., 1997). In stress conditions, proteins may unfold and aggregate. Holding chaperone IbpB binds to partially folded proteins until folding chaperones become available and can intercalate into large aggregates. The holding chaperones Hsp31 and Hsp33 function under severe thermal and oxidative stress, respectively. ClpB promotes disaggregation and in cooperation with DnaK-DnaJ-GrpE refolds proteins, when stress has abated.

During high-level synthesis, folding helpers can be rapidly titrated out and if the protein aggregation is faster than folding, recombinant proteins accumulate in inclusion bodies (IB). A second reason to IB formation is that the formation of disulfide bonds needed for proper folding (e.g. of Fab antibody fragments) is not possible in the reducing cytoplasm of wild-type *E. coli*. IBs can be seen by optical microscope as refractile particles of up to 2 μm^3 (Carrió et al., 1998). IBs have been often regarded as inert and final aggregates, where proteins are inactive and protected from degradation. However, it was shown that arrest of recombinant protein synthesis resulted in efficient removal and refolding of IBs. Although most protein was degraded by proteases, a significant fraction (estimated to be around 25%) gained biological activity, indicating that it had reached native conformation (Carrió and Villaverde 2001). Thus, protein aggregation in bacteria is an active and reversible process that stores misfolded polypeptides until they could be further processed: either refolded or proteolysed. It was shown that proteins in inclusion bodies are not completely inactivated and even the use of IBs as active catalysts without previous refolding step can be considered (García-Fruitós et al., 2005).

Two opposite strategies are successfully used for cytosolic recombinant protein production: preventing or promoting of IB formation.

Production of soluble proteins. Intracellular production in soluble form could be considered if the protein is stable against proteolysis and not harmful to the host cell. A model for the calculation of solubility percentages of recombinant proteins expressed in the *E. coli* cytoplasm (Wilkinson and Harrison, 1991) is available at http://www.biotech.ou.edu.

There are several approaches to enhance solubility of recombinant proteins (Jonasson et al., 2002; Sørensen and Mortensen, 2005): (1) Reducing the expression rate in several ways: by reducing the temperature to 20º C and less during the induction phase is an effective and often used method; by lowering the copy number, using the weaker promoters or lower gratuitous inducer concentrations, e.g. 10-100 μM IPTG (graded induction); by lowering postinduction feeding rate; or by using different substrates. (2) Modifying the gene so that it does not alter desired function or activity. (3) Using highly soluble fusion

partners. (4) Coexpression of chaperones has been used with various success. A number of plasmids compatible with ColE1-derived expression vectors are available (Baneyx and Palumbo, 2003). Chaperones must be produced in large quantities, which is an additional burden to the cell. (5) Using mutant *trx/ gor* strains for proteins with disulfide bonds. (6) Coexpression of binding partners needed for correct folding (expression of whole proteins and complexes instead of their parts). Expression of whole protein sequence (including the pre- and prepro-sequences of secretory proteins and some enzymes) is often necessary to achieve correct structure and functionality. (7) Using *E. coli* mutant strains C41(DE3) and C43(DE3) that allow over-expression of globular and membrane proteins difficult to express (Miroux and Walker, 1996), commercialized by Avidis (http://www.avidis.fr).

It should be also mentioned that soluble proteins do not necessarily have correctly folded structure and activity (Nishihara et al., 2000).

Production as inclusion bodies is reviewed in (Singh and Panda, 2005; Fahnert et al., 2004). For many high-value proteins (e.g. pharmaceutical proteins) IB formation and subsequent denaturation/ refolding into an active form is employed as a first purification step. IBs provide also some protection against proteolytic degradation by host proteases. For lower-value products, as technical enzymes, this strategy could be too expensive. Measures opposite to above described for improvement of solubility are used to promote IB formation, e.g. rising the growth temperature, rising the expression rate by using high copy number or runaway plasmids, strong promoters, fusion to highly insoluble proteins.

Methionine processing (Altenbuchner and Mattes, 2005). In *E. coli* cytoplasm N-terminal initial formylmethionine is removed by methionine aminopeptidase (*map*). Efficient processing depends on the amino acid adjacent to methionine (cleaved in case of Ala, Gly, Pro, Ser; variable result in case of Ile, Val, Cys, Thr; retained in other cases) (Ben-Bassat, 1991). Thus, N-terminal product heterogeneity due to partial processing can occur. Enhancement of processing has been achieved by concurrent expression of *map* gene (Sandman et al., 1995).

Disulfide bond formation (Altenbuchner and Mattes, 2005). Stably disulfide bonded proteins can be found only in the cell envelope of *E. coli* (Baneyx and Mujacic, 2004). Disulfide bonds can be formed by periplasmic oxidoreductases DsbA and DsbB, and rearranged in incorrectly paired cysteins by periplasmic disulfide bond isomerases DsbC and DsbC. Cytoplasmic proteins do not generally contain structural disulfide bonds, although certain cytoplasmic enzymes form such bonds as part of their catalytic cycles. The disulfide bonds in these latter enzymes are reduced by thioredoxin and glutathione/ glutaredoxin pathways (Stewart et al., 1998). Oxidized forms of thioredoxins and glutaredoxins are able to oxidize proteins, but they are kept in a reduced state by thioredoxin reductase (TrxB) and glutathione, which in turn is reduced by glutathione oxidoreductase (Gor). In double mutant *trxB* and *gor* strains, many proteins needing disulfide bonds to achieve native structure could be expressed in cytoplasm in higher amounts than in periplasm. The amount of active protein could be further increased by intracellular

production of the periplasmic disulfide isomerase DsbC (Bessette et al., 1999). *TrxB* and *trxB/ gor* strains are commercially available (AD494 and Origami, Novagen).

Secretory and extracellular production

Secretory and extracellular production is reviewed in (Choi and Lee, 2004; Mergulhão et al., 2005; Baneyx and Mujacic, 2004; Georgiou and Segatori, 2005).

Secretory production. Secretion could be efficient if the protein is exported by its original host cells. Usually, large cytoplasmic proteins cannot be exported even when they are fused to a signal peptide. Most periplasmic and outer membrane proteins in *E. coli* are exported across the inner membrane by the general secretory pathway (Sec) consisting of membrane embedded complex of the SecYEGA proteins. The protein to be exported should have loosely folded conformation and an N-terminal 18 - 30 amino acids long signal peptide, which is cleaved by the signal peptidases during translocation of the polypeptide. Folding of the polypeptide in the cytoplasm is inhibited by binding of SecB and other cytoplasmic chaperones. For efficient secretion, signal sequences from *E. coli* proteins like outer membrane protein OmpT or OmpA, from β-lactamases or form the pectate lyase PelB of *Erwinia carotovora* are often used. Amino acid sequences of common signal peptides can be found in (Choi and Lee, 2004). A second secretion mechanism called twin arginine translocation (TAT) system transports into periplasm proteins that are already folded in the cytoplasm, often with a bound cofactor. TAT signals are similar to Sec signals but have two consecutive arginine residues in the conserved N-terminal part of the signal peptide.

Advantages of periplasmic production are: (1) considerably lower amount of contaminating proteins in the starting material for purification, (2) authentic N-terminus is generated by the signal peptidase, (3) decreased proteolysis due to fewer proteases which are present in lower abundance, (4) simplified protein release by osmotic shock procedures, (5) disulfide bonds can be formed, (6) secreted proteins can be used to *in vivo* enzyme activity assays and also to screen protein libraries (refs. in Choi and Lee, 2004).

On the other hand, secretion of a protein into the periplasm is a bottleneck for high-level production of recombinant proteins. In many cases, preproteins are not readily exported, form IBs or are degraded in the cytoplasm. Incomplete processing of signal sequences can occur and stopping into membrane during translocation can lead to lethality and cell lysis. Recombinant proteins emerging in periplasm in a partially folded form have also to reach a native conformation.

Similar measures as for obtaining soluble proteins in cytoplasm can be used to solve aforementioned problems, e.g. reducing expression rate (lower synthesis rate during longer expression may still provide significant product accumulation while avoiding saturation of the folding and/ or secretory apparatus of the cell), decreasing temperature, overexpressing periplasmic chaperones (e.g. Skp and the PPIase FkpA) and foldases (DsbABCD), and using protease mutant host strains (see Proteolysis).

Extracellular production. In general, laboratory strains of *E. coli* do not secrete proteins into the medium, whereas pathogenic *E. coli* and many other Gram-negative bacteria segregate toxins, proteases, pullulanase and filamentous phages (Altenbuchner and Mattes, 2005). Small proteins secreted into the periplasm are frequently leaked into the culture medium (Tong et al., 2000). There are two ways to direct proteins into the medium (reviewed by Shokri et al., 2003). The first is to direct the target proteins into the periplasm, followed by subsequent passive passage through the outer membrane. Outer membrane can be made leaky using osmotic shock (Neu and Heppel, 1965), incubation with polyethylene glycol, glycine (~2% w/ v), EDTA or Triton X-100, after the cells have been harvested. Leaky L-form cells, wall-less or wall-deficient cells can be used. The outer membrane could also be permeabilised by coproduction of colicin lysis proteins (which deliver the toxin to the medium, e.g. colicin E1 lysis protein Kil). However, the leaky outer membrane can rapidly lead to cell lysis and death. Modified bacteriocin release protein BRP allows can generate permeable zones in the cell envelope (through which protein can pass into the culture medium) without growth inhibition, lysis and contaminating proteins (Van der Wal et al., 1998). A second way is to use an active transport system from another bacterial strain or species. The hemolysin transport system (Hly) forms protein channel between the inner and outer membranes of *E. coli*, and hemolysin is secreted directly from the cytoplasm to the extracellular medium. The target protein to be excreted is fused to the C-terminal hemolysin secretion signal. Active scFv antibodies and human interleukin-6 were released into the medium using hemolysin secretion pathway (Fernández et al., 2000). Proteins could also be fused to carrier proteins that are normally secreted into the medium (e.g. hemolysin) or to a protein expressed on the outer membrane (e.g. OmpF).

Proteolysis

The degradation of misfolded proteins by host proteases guarantees that abnormal polypeptides do not accumulate within the cell and allows amino acid recycling. Proteases have also regulatory functions. In the cytoplasm, proteolytic degradation is initiated by five ATP-dependent heat shock proteases (Lon, ClpYQ/ HslUV, ClpAP, ClpXP and FtsH) and completed by peptidases that hydrolyse sequences 2 - 5 residues in length.

Recombinant gene products are commonly sensitive to proteolysis. In such cases, strains lacking ATP-dependent cytoplasmic proteases may be useful expression hosts, particularly if a single protease is responsible for degradation. However, this approach is not without drawbacks since cell growth and viability might be affected. The fact that *E. coli* B strain BL21 retains good growth characteristics while lacking both *lon* and *omp*T explains the popularity of this host. However, it is not known if the cells compensate for these deficiencies by upregulating the concentrations of other proteases. In periplasm, the DegP is the major housekeeping protease. Misfolded proteins are rapidly degraded by DegP. Strains containing mutations in *deg*P or in other protease genes (*omp*T, *ptr*, *prc*)

can be used to avoid periplasmic degradation. Nevertheless, the tradeoff associated with their reduced growth rates should be carefully considered (Baneyx and Mujacic, 2004).

Fusion protein technology

Fusion proteins consist of a partner or tag linked to the target protein by a recognition site for a specific protease. Fusion partners can be used (1) to enhance solubility, (2) to improve the expression levels (probably because of mRNA stabilisation), (3) to protect target proteins from proteolysis, (4) as expression reporters (5) as folding reporters, (5) for screening, (6) for solubilising partners, and most often for (7) affinity purification. Affinity tags were reviewed in (Terpe, 2003). Common affinity tags are the polyhistidine tag (His-tag), (see Immobilised metal affinity chromatography - IMAC) and the glutathione S-transferase (GST) tag for purification on glutathione based resins. Common solubility enhancing partners are for example maltose binding protein (MBP), N-utilizing substance A (NusA), *Schistosoma* glutathione transferase (GST), and thioredoxin. The outcome of fusion to a solubility partner is protein specific and is not a universal method for the prevention of inclusion-body formation (Sørensen and Mortensen, 2005).

In order to get an authentic protein, the isolated fusion protein is cleaved from fusion partners *in vitro*. Common proteases used for site-specific proteolysis are: (1) factor Xa and (2) thrombin (both from the eukaryotic blood clotting cascade), (3) enterokinase (more specific than previous two), (4) precision protease (highly specific; Amersham Biosciences), and (5) tobacco etch virus protease (TEV). Co-expression of TEV from a recombinant plasmid can be used as a tool to study cleavage of a recombinant target protein *in vivo*.

Co-overexpression technology

In the case of multidomain proteins or protein complexes, coexpression of components is more advantageous than *in vitro* complex reconstitution from isolated components (Sørensen and Mortensen, 2005). Co-expression often results in increased amounts of properly folded target proteins and protection from proteolytic degradation (Li et al., 1997).

Two general strategies are available, namely co-expression from separate plasmids maintained in the cell simultaneously, or expression of multiple recombinant proteins from a plasmid polycistron. A new system for double cistronic co-expression of up to eight recombinant proteins from four different plasmids has recently been commercialized by Novagen. Each plasmid carries different compatible replication origins as well as four different selectable markers necessary for maintaining the plasmids in the cell (Held et al., 2003).

Examples of recombinant products and processes

There are many examples of recombinant proteins produced in high-cell density culture (HCDC) and by secretion technologies reviewed in (Choi et al., 2006; Choi and Lee, 2004).

One of the most complex proteins expressed in *E. coli* is human tissue plasminogen activator (tPA). Truncated tPA with nine and whole tPA with 17 disulfide bonds were expressed in the active form in cytoplasm (Bessette et al., 1999) and periplasm (Qiu et al., 1998), respectively. These examples demonstrate the potential of *E. coli* as production host.

1.1.8 Strain improvement and expanding the potential of *E. coli*

Host organisms have been genetically modified to improve quality of the recombinant proteins and the yield of production processes. As described in previous sections, hosts can be engineered to ensure plasmid maintenance, to support the translation of rare codons, to reduce proteolytic activities, to facilitate protein folding (to provide more chaperones) and formation of disulfide bridges in cytoplasm, to achieve secretion, etc.

For protein expression, *E. coli* B and K12 strains and their derivatives (e.g. BL21) are most frequently used. BL21 produces less acetate, the inhibitory by-product of growth. Several genetic approaches to reduce acetate formation are described in (Swartz, 2001), but limited glucose feeding is the universal solution for this purpose (see Cultivation).

One example of host engineering is the quiescent cell expression system. The overexpression of a small RNA called Rcd in hns mutant strains (to avoid production of the H-NS histone-like nucleoid structuring protein) produced a quiescent cell state where cells significantly slowed their production of host proteins, but continued to express the recombinant product (Rowe and Summers, 1999). In these cells, cellular resources are focussed on product synthesis and the possibility of secondary induction of proteases is avoided.
Another such example is the RNA integrase expression system (Inouye, 2006).

Glycosylation in E. coli*?*

The three main disadvantages of *E. coli* expression system are considered to be: (1) inefficient folding and formation of disulfide bridges, (2) poor secretion, and (3) lack of glycosylation capability. Recently it has become apparent that the last of these may also be addressed by additional engineering of *E. coli*. N-linked glycosylation of a model protein was achieved by glycosylation pathway of *Campylobacter jejuni* transferred to *E. coli* (Wacker et al., 2002). Another strategy for glycosylation was developed by (Zhang et al., 2004), the co-translational synthesis of selectively glycosylated proteins in which the modified amino acid is genetically encoded.

1.2 Cultivation

High cell density cultivation (HCDC) is reviewed in Choi et al., 2006; Shiloach and Fass, 2005; Lee, 1996. Clever genetic solutions (like those used for enhancing plasmid stability or for reducing acetate production) cause often additional burden and/ or impairing essential functions, but are useful tools to study such phenomena.

The yield and quality of recombinant protein depends upon the characteristics of the gene-vector-host combination as well as on cultivation of producer cells, and finally, on downstream processing. Important issues in cultivation are: media composition, feeding strategies and cultivation variables (e.g. temperature, dissolved oxygen level (DO) and pH). All of these can affect transcription, translation, proteolytic activity, secretion, expression levels and stability, and have to be optimised to achieve high production yields (Choi et al., 2006).

In the case of intracellular production, the amount of recombinant protein produced is proportional to the amount formed per cell and the final cell density. HCDC techniques not only improve productivity, but also reduce culture volume, enhance downstream processing, reduce wastewater, decrease production costs and reduce investment in equipment (Lee, 1996).

Densities from about 20 up to 100 and more grams dry cell weight per litre have been reported as HCDC (for comparison, 1 g dw l^{-1} ~ $1.7 \cdot OD_{600}$). In practice, densities significantly higher than those achievable in batch culture can be considered as HCD.

During the last three decades many research groups have developed growth media, aeration techniques, monitoring devices, controlled fermentation systems, and culture strategies needed for achieving high density cultures. Roughly, these techniques were developed and designed to establish growth conditions that keep the majority of cells in a growing *E. coli* culture satisfied and happy (Shiloach and Fass, 2005).

1.2.1 Cultivation media

Media can be complex, semi-defined or defined. Standard medium for small-scale cultures is Luria-Bertani (LB) complex medium (Sambrook et al., 1989). The small-scale batch (see Cultivation modes) cultures are grown in test tubes and conical flasks on the shaker (to provide aeration), and are used for screening of recombinant cultures, for expression of recombinant proteins and for fermentation inoculum preparation. On LB medium OD_{600} values reach up to 2 - 6 and growth stops due to nutrient and energy limitations, and pH rises up to 8.8 due to the accumulation of ammonia. An additional energy source (e.g. glucose or glycerol) avoids the rise in pH, but inhibitory by-products start to accumulate due to the exhaustion of respiratory capacity and overflow metabolism (Paalme et al., 1990, 1995). At biomass concentrations higher than 0.5 g dw l^{-1} the conditions for production of recombinant proteins become unfavourable.

Defined media contain mineral salts, a carbon source (usually glucose), and other supplements (e.g. amino acids, vitamins) if needed. These media, for example the minimal medium M9 (Sambrook et al., 1989), can be more suitable for inoculum preparation and even for screening of cultures for production. This is because the carbon sources and salts can be the same at small-scale and in the high density fermenter cultures, and therefore the behaviour of the culture, especially the expression stability is more similar on both scales (our observations).

For large-scale cultivation in a fermenter, defined media are generally used, as the concentrations of the nutrients are known and feeding can be controlled during the cultivation. To obtain HCD, a balanced medium is necessary that contains all the nutrients needed for growth, while avoiding inhibition. A balanced medium can be designed based on the elemental composition (C: N: P: S, microelements) of *E. coli* dry biomass and the yield coefficients of the medium components (Paalme et al., 1990; Riesenberg et al., 1991). It is reasonable to design the starting media so that it contains all mineral components except ammonia required to obtain final cell density and only a small initial amount of carbon source, whereas a feeding medium is a concentrated solution of carbon source and additional salts. Ammonium hydroxide is frequently used as both a nitrogen source and pH control agent (Lee, 1996).

Carbon sources

E. coli can grow on a variety of single carbon sources and works to convert these to biomass, carbon dioxide, water, and, depending on the conditions, acetate and other organic acids. The monomers from which the cell's polymers are made are synthesised from 12 precursors that are the products of the central metabolic pathways (CMPs) (Neidhardt et al., 1990). All the precursors must be provided to generate new biomass, which is achieved by minor variations of the active CMPs and specific pathways to take up the particular carbon source and deliver it into the CMPs. Most often, glucose is used as a carbon source, which is also the preferred substrate for *E. coli*.

Glucose is taken up and phosphorylated by the phosphotransferase system (PTS), which simultaneously generates pyruvate from phosphoenolpyruvate (PEP). Glycerol is taken up by facilitated diffusion across the cytoplasmic membrane through the action of the facilitator protein GlpF and is converted into triose phosphates through the glycerol dehydrogenase reaction, giving glucose 6-phosphate (G6P) through gluconeogenesis but can feed into the pentose phosphate pathway (PPP) from triose phosphate (C3P) without oxidative decarboxylation. Acetate is taken up by a proton symport, phosphorylated by acetokinase to acetyl phosphate and thence to acetyl-CoA by a phosphotransacetylase. If acetate is the sole carbon source, the synthesis of all the other precursors depends on the glyoxylate bypass, which generates malate and thence oxalacetate (OAA) from acetyl-CoA by isocitrate lyase and malate synthase. Succinate is taken up by the C4-dicarboxylate transporter DctA (Davies et al., 1999) and is a TCA cycle intermediate. For growth on so-called gluconeogenic substrates like glycerol, the succinate and acetate gluconeogenesis reactions are activated, which are essential for the formation of glucose from nonhexose precursors. Pyruvate is converted to PEP by PEP synthetase, and fructose 1,6-diphosphate to fructose 6-phosphate through fructose biphosphatase. Conversion of oxaloacetate directly to PEP through PEP carboxykinase is also possible (Phue et al., 2005).

An alternative to glucose carbon sources can be used to reduce the growth rate, to avoid by-product formation or to reduce the expenses of labelling.

Semi-defined media and addition of amino acid(s)

Occasionally, semi-defined media (defined media supplemented with complex additives, such as peptone, yeast extract, casamino acids) are used to boost growth and/ or recombinant product formation. However, the common practice of adding high levels of supplemental amino acids may inhibit recombinant protein production due to feedback inhibition of critical biosynthetic pathways (Harcum, 2002). Therefore, increased recombinant protein yields could be obtained only by a well-designed amino acid addition strategy, for example, if predicted amino acid shortages are responsible for the stringent response and associated upregulated protease activity (Harcum and Bentley, 1999).

1.2.2 Growth rates

For *E. coli* K12 strain W3350 the highest specific growth rate obtained in fermentations with glucose and casamino acids is 0.77 h^{-1}. In minimal medium, the maximal specific growth rates at 37°C are 0.57 h^{-1} with glucose and 0.32 h^{-1} with glycerol, determined by continuous accelerostat (A-stat) cultivation (Paalme et al., 1997). Growth rates on glyconeogenetic carbon sources are lower than on glucose, however the exact values are strain dependent.

1.2.3 Cultivation conditions

Temperature

Escherichia coli is a mesophile, for which balanced growth can be sustained from 10ºC to almost 49ºC. In the so-called normal or Arrhenius range (20 to 37°C) the specific growth rate of *E. coli* increases with temperature without a considerable change in protein levels as though cellular growth acted as a simple chemical process (Herendeen et al., 1979). Above and below this range the growth rate does not follow this relationship (remaining practically constant from 37 to 42ºC) and heat shock and cold shock responses are induced, respectively (Farewell and Neidhardt, 1998; Wick and Egli 2004), until growth ceases. Usually, *E. coli* is cultivated at 37ºC, i.e. on the edge of a normal range. However, lower temperatures often give better stability for the production of recombinant cultures, and therefore give better yields and or quality of recombinant proteins (Friehs, 2004; our results).

pH

E. coli prefer neutral pH, but normally grow over the range of pH 5 to 9, for which internal pH is regulated at pH 7.4 to 7.9 during growth (Small et al., 1994). During *E. coli* fermentation, pH decrease is usually avoided by titration with ammonium chloride solution.

Oxygen

It is very important to maintain aerobic conditions throughout the cultivation and recombinant protein synthesis. This can be achieved by increasing the agitation

speed, the aeration rate and by supplying oxygen enriched air or pure oxygen, if needed. In the case of recombinant cultures, dissolved oxygen (DO) is usually kept above 20% of air saturation. The oxygen consumption of the culture can be reduced by lowering the growth rate (by using alternative carbon sources, lower temperatures and limitation of carbon source).

1.2.4 Acetate problem

Overcoming the formation of acetate in cultivation was recently reviewed (Eiteman and Altman, 2006). It is generally accepted, that acetate starts to inhibit *E. coli* from the concentration of 2 g l^{-1} (Shiloach and Fass, 2005), but in case of K12 ATCC 2732 and BL21 strains acetate retards growth even at concentrations as low as 0.5 - 1 g l^{-1} (Nakano et al., 1997; our results, respectively). Acetate also inhibits recombinant protein synthesis and interferes with many other synthesis and regulation processes (Eiteman and Altman, 2006).

E. coli cells produce acetic acid as an extracellular co-product at high specific growth rate when there is excess glucose or under oxygen-limited conditions. Despite fully aerobic conditions, *E. coli* can consume oxygen only up to a maximum rate, which corresponds to the critical growth rate at which acetate formation begins (Paalme et al., 1997). The critical specific growth rate that leads to acetate formation varies among strains and depends on the medium used (Lee, 1996). The formation of acetic acid also depends on the carbon source (Holms, 1996). Glyoxylate and glyconeogenesis pathways essential for acetate utilisation are constitutively active in *E. coli* BL21, and as a result BL21 accumulates considerably less extracellular acetate than does *E. coli* K strain JM109 although both produce acetate at the same specific rate (Van de Walle and Shiloach, 1998; Phue et al., 2005).

There are several ways to overcome acetate (Eiteman and Altman, 2006): (1) using alternative carbon sources e.g. glycerol. During growth on glycerol, the flux through glycolysis decreases and, depending on the strain, less or no acetate is produced (Chung et al., 2000; Korz et al., 1995; Lee 1996; Holms, 1996; Luo et al., 2006); (2) by supplementing medium with glycine, methionine, yeast extract or amino acids (Han et al., 1993); (3) lowering the cultivation temperature (e.g. Horn, 1996); (4) by metabolic engineering of *E. coli* (This approach has been much studied recently); (5) by using a dialysis culture, however, nutrients are wasted in this case; (6) by using strains that produce less acetate (e.g. RV308, BL21); (7) by controlling the growth rate by limiting the carbon source (i.e. using feeding strategies). If another nutrient limits growth (e.g. nitrogen), then *E. coli* invariably accumulates acetate even at low growth rates because the threshold critical glucose consumption rate at which acetate formation starts, is at its maximum value when glucose is in excess (Hua et al., 2004).

1.2.5 Cultivation modes

Batch, fed-batch and continuous cultivation modes have been used with *E. coli*. Batch is the most often used and simplest mode of cultivation. One additional

advantage is that batch cultures can be carried out in small volume, which is important in multiple parallel experiments or in case of expensive, for example labelled media. In batch, all the nutrients are given in the medium at the beginning and growth conditions are constantly changing with time. Due to growth, unfavourable changes occur in the cultivation medium, such as oxygen limitation, pH changes, accumulation of inhibitory by-products and consumption of the essential substrates. Usually, only moderate cell densities and volumetric production levels can be obtained in batch cultures (Lee 1996). This is true even for a fermenter, where DO and pH can be kept constant. In addition, the quality of recombinant protein can be a problem.

Continuous mode, e.g. chemostat cultivation (Monod, 1950; Novick and Szilard, 1950) is used mainly for studying the microbial physiology, because reproducible steady state conditions can be achieved and maintained for long periods. In chemostat culture feeding medium is fed at a constant rate and culture volume is kept constant. An efficient continuous cultivation method, where dilution rate is changed smoothly, was developed for quantitative study of cell physiology (Paalme et al., 1992; Paalme et al., 1995) and is commonly used in industrial practice for determination of critical growth rates. However, recombinant bacteria are usually not cultivated in continuous cultures, because this mode is ineffective for the production of inducible proteins, and also because non-producing cells can overgrow due to the longer runs.

Fed-batch processes have most often been used to obtain high cell density (Yee and Blanch, 1992; Riesenberg and Guthke, 1999; Lee et al., 1999), which is one important prerequisite to achieve high volumetric productivity of recombinant proteins. Fed-batch cultivation starts in batch mode, where bacteria grow fast up to optimal concentrations prior to feeding. Upon feeding, various nutrient feeding strategies are applied to achieve high density. To avoid by-product formation, it is advantageous to start the fed-batch cultivation just before the carbon source becomes consumed.

1.2.6 Feeding strategies of fed-batch cultivation

Feeding methods affect the growth, attainable biomass concentration, and the yield and quality of recombinant protein. Fed-batch cultivations are normally carried out under carbon limiting conditions.

Feeding strategies with feed-forward control

Predetermined (e.g. constant, increased or exponential) feeding rate can be used for obtaining HCD.

μ-stat cultivation. The exponential feeding algorithm (Paalme et al., 1990; Lee, 1996) allowing the cells to grow at constant specific growth rate *(μ)* is called μ-stat. Respective feeding rate can be calculated using the equation derived from mass balance (Lee, 1996):

$$M_S(t) = F(t)S_F(t) = (\frac{\mu}{Y_{X/S}} + m)X(t)V(t) = (\frac{\mu}{Y_{X/S}} + m)X(t_0)V(t_0)\exp[(t - t_0)]$$

Where: M_s - mass-flow rate of the carbon source (g h^{-1}); F-feed flow rate (h^{-1}); S_F - carbon substrate concentration in the feed (g l^{-1}); X - cell concentration (g l^{-1} dw); m - specific maintenance coefficient (g g^{-1} dw h^{-1}); V - culture volume (l); t_0 - time of feeding start; t - process time; μ - specific growth rate (h^{-1}); $Y_{X/S}$ - cell yield on carbon substrate (g g^{-1}), which is assumed to be constant.

Feeding strategies with feedback control

Direct feedback control. Nutrient feeding is directly controlled by online measurement of the concentration of the carbon source. For glucose, the method does not work for *E. coli* because K_S is too low (in the range of 40 - 88 μg l^{-1}) (Senn et al., 1994).

Indirect feedback control. This method is based on online monitoring of pH, DO, CO_2 evolution rate or cell concentration (e.g. OD). When the pH or DO begins to rise due to the carbon source depletion, a predetermined amount of concentrated feeding solution is pumped into the fermenter. In a defined medium, the DO-stat responds more rapidly to nutrient depletion than pH-stat. pH-stat in turn is more suited to growth on semi-defined media (Lee, 1996).

Ethanol sensors are used as feedback control in the industrial cultivation of yeast to prevent overflow metabolism (ethanol formation). Suitable acetate sensors to control the overflow metabolism in *E. coli* are not available.

Probing control method. Under glucose-limited conditions, superimposed glucose pulses are applied to the feed. A change in DO-response indicates that growth rate is less than critical to cause the overflow metabolism. Otherwise there is no or insignificant response. (Indirect) on-line detection of acetate formation makes it possible to avoid overflow metabolism using feedback control of the glucose feed rate. A simple feedback algorithm using the information from the pulse responses has been developed to achieve feeding around the highest glucose uptake without acetate formation (Åkesson et al., 2001).

Adaptastat method adapts cultures to their maximum feeding rates while avoiding accumulation of substrate(s) and by-products (V). The latter feedback methods allow to avoid overfeeding (and acetate production) even in case of inevitable process and metabolic disturbances (including those caused by induction of protein expression).

1.2.7 Induction conditions

Induction temperature

Although recombinant *E. coli* may grow fastest at 37ºC, the optimal temperature for heterologuous gene expression is often lower, typically between 25 - 30ºC, or even lower. In some fermentation strategies a lower temperature is also applied

during the growth phase to minimize stress prior to the production phase, especially when using strong promoters (Altenbuchner and Mattes, 2005).

Duration of induction

Induction times in the range of 4 - 40 h have been reported. Accumulation of a given recombinant protein is determined by the specific speed of its formation and degradation, and is protein-specific. It has been shown in the case of some very strong promoters (e.g. the T7 promoter) that higher productivities can be obtained when applying a down-modulated induction during more sustained prolonged induction period. In addition, a feed of the same amount of IPTG compared to single-dose addition could be better in order to adapt the cells to production (Sandén et al, 2003).

Pre- and post-inducion feeding

In case of inducible expression systems, fed-batch cultivation process can be divided into growth and production phases. The culture is induced when the cell density reaches a suitable value. It is reasonable to maintain the pre-induction feeding rate just below the critical value to avoid the accumulation of inhibitory by-products (mainly acetate) while at the same time ensure high RNAP activity and ribosome content in the cell (Sandén et al, 2003; Vanatalu et al., 1993). In addition, at lower growth rates the biomass yield may decrease due to maintenance (Pirt, 1965). Recombinant protein production can be significantly affected also by post-induction feeding strategies (Wong et al., 1998; III; our unpublished results). Both, insufficient feeding and overfeeding can result in lower protein production. After induction the cell growth and substrate consumption rate usually decrease significantly. Adequate feeding without overfeeding can be provided via feedback control using the adaptastat method (V).

Not only productivity, but also product quality is important for the production of recombinant proteins. Very few reports are currently available on the effect of the feeding profile on the quality of the product (Sandén et al., 2003). It was shown by Ryan et al. (1996), that a high productivity was inversely proportional to a high product quality and more isoforms of the protein were present the higher the production rate. Isoforms of a protein are synthesised due to mistranslation, which could misincorporate or omit amino acids (see Codon usage). Thus, high (or too low) synthesis rates of heterologous proteins with different codon usage might cause lower quality. Hellmuth et al. (1994) showed that there is an optimum in the achieved aprotinin-β-galactosidase fusion protein activity as a function of the specific growth rate.

The quality of rat COMT, purified from biomass obtained by fed-batch procedure (III) was better in comparison with that from conventional low-density high-volume bath culture on rich medium (Dr. Ismo Ulmanen Orion Pharmaceutica Target Protein Laboratory, personal communication).

1.3 Labelling of recombinant proteins for 3D structure determination

1.3.1 Protein NMR

The three-dimensional structure of proteins is required to elucidate their function. Most atomic resolution structures of proteins have been solved either by X-ray diffraction or by nuclear magnetic resonance (NMR), and are deposited in the Protein Data Bank (Berman et al., 2003; http://www.rcsb.org/pdb). NMR profiling of proteins was recently reviewed (Gardner and Kay, 1998; Wider, 2000; Rule and Hitchens, 2006). For X-ray diffraction, proteins have to be crystallized, whereas NMR experiments can be carried out under near physiological conditions in solution. NMR structures can be obtained also for proteins that cannot be crystallised. In addition, NMR is useful in studying molecular interactions (e.g. protein-protein interactions, complex formation).

NMR does not directly produce an image of a protein, but gives ample indirect structural information from which the structure can be deduced by complex analysis and calculation (Wider, 2000). NMR signals can be obtained from atomic nuclei with spin of 1/2 that are magnetically active. In proteins these are nuclei from hydrogen (^{1}H), and from less common stable (non-radioactive) isotopes of carbon (^{13}C) and nitrogen (^{15}N). ^{13}C and ^{15}N have the natural abundance of 1.1% and 0.37%, respectively. Stable hydrogen isotope deuterium (^{2}H, 0.0155%) plays also a very important role in NMR. When applying radio frequency electromagnetic radiation to the protein sample in high magnetic field, every NMR-active nucleus gives rise to an individual signal (resonance line in the NMR spectrum). The resonance frequency depends on the nucleus type (e.g. ^{1}H, ^{13}C) and on the chemical environment of that nucleus (i.e. on the interactions with surrounding nuclei and electrons). As a result, information about the nucleus and chemical environment can be derived from its resonance frequency. The structure determination is based on the fact that resonances can be assigned to every nucleus by NMR techniques.

NMR spectra of proteins contain high number of resonances whose assigning becomes more difficult for larger proteins. The study of proteins by NMR has greatly benefited from the use of isotope labelling (reviewed in Lian and Middleton, 2001; Goto and Kay, 2000).

1.3.2 Isotopic labelling in protein NMR

^{15}N- and/ or ^{13}C-labelling

The solving of proton (nucleus of ^{1}H) NMR spectra for proteins larger than about 10 kDa is prohibitively complicated due to signal overlapping and broad signals in crowded spectra. To overcome these difficulties, uniform labelling of proteins with ^{15}N and/ or ^{13}C, along with heteronuclear and multidimensional NMR experiments was developed. 2D, 3D and 4D NMR spectra provide increased resolution and correlations between different nuclei that results in spectra that are easier to

analyse. This allows one to determine the structure of small and medium sized proteins up to 35 kDa. Most of these experiments use three types of nuclei, ^{1}H, ^{15}N and ^{13}C, and are referred to as triple resonance experiments (Wider, 2000).

^{2}H-labelling

In addition to signal overlapping, sensitivity is low due to faster relaxation with increasing molecular weight. However, ^{15}N- and/ or ^{13}C-labelling have little effect on nuclear relaxation and rather increase line widths. The replacement of hydrogens by deuterium (^{2}H), amongst all the isotope labelling strategies, has contributed most to improved sensitivity and resolution (Lian and Middleton, 2001). Deuteration improves the resolution and sensitivity of NMR experiments by reducing the overall relaxation rates of NMR-active nuclei, in particular of ^{13}C nuclei because the 6.7 fold lower gyromagnetic ratio of ^{2}H relative to ^{1}H (Leiting et al., 1998). ^{2}H has a different resonance frequency, is less sensitive and yields broader signals than ^{1}H. Reduction in the number of peaks and concomitant narrowing of line widths greatly simplifies the spectra (Gardner and Kay, 1998).

The accessible molecular weight range can be extended to about 50 kDa by fully deuterated $^{13}C^{15}N$-labelled proteins (Wider, 2000). Experiments with ^{2}H/ ^{13}C/ ^{15}N-labelled proteins facilitate sequence-specific resonance assignment of the backbone atoms and the side chain heavy atoms. However, it is still essential to assign side-chain protons because the through-space correlations between protons provide most of the spatial information used to calculate the structures. Thus, in addition to fully deuterated proteins used for sequential (through-bond) assignment, a partially (50-70%) deuterated protein is required to collect structural information from through-space proton-proton correlations (Wider, 2000). The degree of deuteration needed depends on the size of the protein, spectral quality and type of NMR experiments used (Goto and Kay, 2000).

Even larger (>40 kDa) proteins and complexes can be studied using advanced NMR and labelling techniques. Relaxation in large molecules could be reduced using transverse relaxation-optimised spectroscopy (TROSY, Pervushin et al., 1997), but also spectral crowding can be reduced by segmental labelling of protein (Yu, 1999). In addition to uniform labelling, reviewed and described in this work, a variety of selective labelling approaches (e.g. site specific and amino acid selective labelling) are used to improve and simplify the NMR spectra (Goto and Kay, 2000; Lian and Middleton, 2001).

1.3.3 Production of labelled proteins

For the production of labelled recombinant proteins all the topics discussed so far in the literature review are relevant. In addition there are specific problems, namely the high price of labelled substrates, including heavy water, and in case of deuterium labelling, the effect of ^{2}H on biological processes.

Media and substrates used for labelling

For isotopic labelling, modified minimal media, e.g. M9 (Sambrook et al., 1989) are commonly used. For nitrogen, carbon, and deuterium labelling media containing ^{15}N ammonium chloride/ sulphate (Muchmore et al., 1989), ^{13}C-glucose (e.g. Cai et al., 1998) and deuterium oxide are mostly used as sole isotope sources, respectively. For perdeuteration, i.e. for replacing all the nonexchangeable carbon-bonded hydrogens, ^{2}H-glucose is commonly used in addition to $^{2}H_2O$. To improve the growth and expression, media have often been supplemented with trace elements and vitamins, commercially available isotopically enriched algal and microbial hydrolysates, or alternatively, commercial labelled media have been used (Cai et al., 1998; Marley et al., 2001; Reilly and Fairbrother, 1994).

Production of labelled proteins is expensive, especially if uniform double or triple labelling is needed and/ or expression level is low. Different strategies have been used to improve the yield per gram of labelled substrate and reduce the labelling costs: (1) In order to minimize the wasting of label into the biomass, the cells are cultivated in unlabelled medium first and labelled nutrients are provided only shortly before the induction. This method has been used in shake flasks (Marley et al., 2001) and in fermenters (Cai et al., 1998; Fiedler et al., 2007). However, this method necessitates very low basal expression level (which is seldom the case), and labelling rates near 100% cannot be achieved. In the fermenter, changing of unlabelled substrate with labelled one can be done when the former is depleted and unfavourable centrifugation and washing stages, needed in case of shake flask cultures, can be avoided. In addition, the better growth conditions can be provided (aeration and pH). (2) Less expensive labelled substrates have been employed for labelling, like acetate (Venters et al, 1991; Venters et al., 1995). To improve the growth, succinate and acetate together have been also used (Vanatalu et al., 1993; IV; this work). (3) High cell density fermentation has been used (Vanatalu et al., 1993; IV; this work), which is most effective in the case of deuteration because the heavy water expenditures can be significantly reduced.

Adaptation to growth in deuterated medium

Recombinant cells can grow and express normally in ^{13}C- and/ or ^{15}N-labelled media. *E. coli* can grow on ^{2}H-media, but growth rate, biomass and recombinant protein yields are usually lower. Because of the inhibitory effect of $^{2}H_2O$, growth and expression rates are lower in partially or fully deuterated medium (Patzelt et al., 2002). As discussed above, the required extent of deuteration depends on the type of experiments that will be performed. To achieve high levels of deuteration (>90%), the modification of culture media, growth protocol and an adaptation procedure, where cells are grown in media with successively rising D_2O percentage may be necessary (Lian and Middleton, 2001). Adapted cells can be stored as glycerol stocks for later use, but expression can be lost in recombinant cells. Alternatively, direct transferring from H_2O to >90% $^{2}H_2O$ can be tested (Venters et al., 1995).

1.4 Purification

Recombinant proteins can be produced in three compartments of *E. coli* in soluble or inclusion body forms, from which they follow different purification processes (Choi et al., 2006). Purification of recombinant proteins is reviewed also in (Jonasson et al., 2002; Swartz, 2001).

His-tagging allows to obtain highly purified recombinant proteins in one step immobilised metal affinity chromatography (IMAC). Differently from other common fusion partners/ affinity tags, his-tag allows purification also in fully denaturating conditions, which is important in purification of insoluble aggregated or membrane-bound proteins, or proteins from non-lysed cells. Histidine-tagged labelled fusion proteins in both, IMAC in native and denaturating conditions were purified in this study.

1.5 Motilin

Motilin is a gastrointestinal peptide hormone, which acts, as many peptide hormones through membrane-bound G-protein coupled receptors. It regulates the contractions of the gastrointestinal smooth muscle. Motilin has been shown to modulate also the action of certain neural tissues (for a review see Poitras, 1994). A number of potential clinical and biological applications have stimulated a large interest in the basic structure-function relationships for motilin. Human (and porcine) motilin (22 amino acid residues; 2699 Da) has the sequence Phe-Val-Pro-Ile-Phe-Thr-Tyr-Gly-Glu-Leu-Gln-Arg-Met-Gln-Glu-Lys-Glu-Arg-Asn-Lys-Gly-Gln.

$^{13}C/\ ^{2}H/\ ^{15}N$-labelled motilins were produced in 3 - 10 mg amounts for structural studies by high resolution NMR spectroscopy (VII, this thesis).

THE AIM OF THE STUDY

The aim of this study was to develop methods for the production of $^{13}C/\ ^{2}H/\ ^{15}N$-labelled recombinant proteins/ peptides for structural studies.

This thesis is based on the following studies:

1. Stability studies of the expression systems (I, II).
2. Development of fed-batch methods for high cell density (HCD) expression – feed-forward (III) and adapatastat (V).
3. Development of methods for cultivation of BL21(DE3) in deuterated medium (VI)
4. The application of methods developed in high cell density cultivation (HCDC) for production of unlabelled and labelled biomasses (III, IV, VII).

The methods developed as well as experience and knowledge obtained in these studies allowed achieving the goal – to produce of highly pure $^{13}C/\ ^{2}H/\ ^{15}N$-labelled motilin for NMR analysis in 1-10 mg amounts. Development of procedures for the production of labelled motilins is described below.

2 MATERIALS AND METHODS

2.1 *E. coli* strains and expression vectors

2.1.1 Strains

Esherichia coli strains BL21(DE3), JM109(DE3), RV308, MRE600rif, MRE600, JM109, and XL1-Blue together with the plasmids and cloned proteins used in this study are listed in Table 1.

2.1.2 Expression vectors

To develop a fed-batch procedure for production of recombinant proteins (III) a number of expression vectors were used and are also listed in Table 1.

To optimise the expression conditions in deuterated media and purification procedure, a recombinant strain expressing the fusion protein GST-GFPuv was used (Table 1). The gene of GFPuv, a mutant form of *Aequorea victoria* GFP (Crameri et al., 1996), was cloned from pGFPuv plasmid (Clontech) into a pET42a(+) expression plasmid (Novagen). The construct was then transformed into a host strain *E. coli* BL21(DE3) (Novagen).

To produce labelled motilin two expression vectors were used: one for ubiquitin human motilin fusion protein (motilin-FP) and another for cleaving enzyme - yeast ubiquitin hydrolase (YUH).

(1) Motilin was cloned using a peptide expression system pTKK19 (pET 19); T7*lac* by Kohno and Sugai (Kohno et al., 1998; VII). The vector contains the gene coding motilin-FP with an N-terminal decahistidine tag (Fig. 1A). Ubiquitin, being a small fusion partner with 76 amino acids, protects motilin (22 amino acids) from proteases.

Table 1. Recombinant proteins, expression vectors and host strains used in the present study

Protein	Fusion tag/ partner	Vector, promoter	Host	Source of construct	Results	Reference
Rat catechol-*O*-methyltransferase		pKEX14, T7	BL21(DE3)		BM	III
Human catechol-*O*-methyltransferase		pKEX14, T7	BL21(DE3)		BM	III
Calf prochymosin		pGEX-2T, T7	JM109(DE3)		BM	III
Human troponin C	GST-	pJEX2, T7	JM109(DE3)		BM	III
Fab		secretable	RV308	M. Karp[1]	HCDC	Unpublished
Sperm whale myoglobin		pMb, constit.	MRE600rif	F. Parak[2]	*^{2}H-BM, 200 g	Tomson and Vanatalu, 1997 (Abstract)
E. coli chaperons GroEL, GroES		pOF39, own promoters, constit.	MRE600; JM109	H. Heumann[3]	*BM	Holzinger et al., 2000; Rössle et al., 2000
E. coli ribosome protein L29	-His	PQE 70, T5*lac*	XL1-Blue	K. Nierhaus, U. Stelzl[4]	*^{15}N-BM	Unpublished
Human vascular endothelial growth factor (VEGF165)		pET; T7	BL21(DE3) BL21(DE3)pLysS	Siemeister et al., 1996	*BM	Unpublished
VEGF121	His-					
Yeast ubiquitin hydrolase (YUH)	-His	pET-20b; T7	BL21(DE3)	Kohno et al., 1998	BM, YUH	VII
Human motilin	His-ubiquitin-	pTKK19 (pET 19); T7lac	BL21(DE3) BL21(DE3)pLysS	VII	*Motilin	VII; this thesis
Aequorea victoria GFP (uv mutant)	GST-	pET42GFPuv T7*lac*	BL21(DE3)	pGFPuv (Clontech)	*BM	Unpublished

[1]Department of Biotechnology, University of Turku; [2] München Technical University; [3]Max-Plank Institute, Martinsried; [4] Max-Planck Institute for Molecular Genetics, Berlin. *Labelled biomasses (BM) were produced, e.g. of *E. coli* MRE600 pOF39 in 100% deuterated medium containing succinic and acetic acids and in 82% deuterated medium containing H-glucose; *E. coli* JM109 pOF39 in 82% and 90% deuterated medium containing H-glucose, all with chaperon proteins GroEL and GroES expressed.

(2) Yeast ubiquitin hydrolase, containing a hexahistidine tag on its C-terminus was cloned into pET-20b; T7. YUH (Fig. 1B) cleaves precisely and specifically behind ubiquitin (Kohno et al., 1998) resulting in a motilin peptide with correct sequence (VII, this thesis). His-tagging is used for protein purification by immobilised metal affinity chromatography (Bornhorst and Falke, 2000).

2.1.3 Transformation of expression vectors into *E. coli* host strains

Expression vectors for motilin-FP and YUH were transformed into *E. coli* BL21(DE3) and BL21(DE3) pLysS (both from Novagen) Ca-competent cells according to Ausubel et al., 1987. The same protocol was used for GST-GFPuv and VEGF constructs.

2.2 Culture media

Complex medium LB and minimal medium M9 (Sambrook et al., 1989) containing different carbon sources and appropriate antibiotics (kanamycin in case of pMot and pYUH at final concentration of 50 μg ml^{-1}) were used for batch cultivations. Yeast extract and Tryptone were obtained from Difco, all other reagents were of analytical grade and obtained from various manufacturers. Labelled reagents used for different types of uniform labelling (Table 2) were obtained from Spectra Stable Isotopes (USA), Cambridge Isotope Laboratories (USA) and $^{2}H_2O$ from Izotop (Russia).

Table 2. Labelled reagents used for different types of uniform labelling

Labelling type	Labelled reagents used	Isotopic purity, % atom of
^{15}N	$^{15}NH_4Cl$	^{15}N >98%
$^{2}H^{15}N$	$^{15}N^{2}H_4Cl$, ^{2}H-succinic acid, ^{2}H-acetic acid, $^{2}H_2O$	^{15}N >98%, ^{2}H >98-99%
$^{13}C^{15}N$	^{13}C-glucose, $^{15}NH_4Cl$	^{13}C >98%, ^{15}N >98%

2.3 Cultivation

2.3.1 Batch cultivation

Batch cultivation in test tubes and shake flasks was used to prepare the inocula and for preliminary optimisation of the growth and expression conditions.

Induced and uninduced "diluted control" cultures: before induction, an aliquot of the fermenter culture was inoculated into M9 medium containing the same substrates (in the case of ^{2}H-labelling the same deuteration percentage was also used) at an initial OD_{600} of 0.5 and grown on the shaker for 3 - 5 h. In each case two experiments were conducted, one with the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) into the medium, and one without.

2.3.2 Fed-batch cultivation

The feed-forward method used is described in (Paalme et al., 1990), the μ-stat algorithm for recombinant expression in III; and the adaptastat method in V. Production of recombinant (labelled) biomasses was carried out in various fermenters under computer control using the software FermExpert (Vinter et al., 1992) with fed-batch cultivation (μ-stat and adaptastat technologies) carried out under carbon limiting conditions. Details are described in III and VII. In some studies continuous cultivation method was also used (II, V).

2.3.3 Induction of recombinant protein synthesis

For induction of recombinant protein synthesis in expression systems, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.5 mM. Batch cultures were induced at OD_{600} of 0.2-0.5, fermenter cultures at various densities.

2.4 Testing of plasmid stability

Samples from the culture were serially diluted with 0.9% NaCl, 0.1 ml aliquots plated onto LB agar (Sambrook et al., 1989) both with and without respective antibiotic, and incubated at 30°C/ 37°C for 24 h. Plasmid stability was also tested by isolation, restriction and agarose gel electrophoresis of plasmid DNA (Sambrook et al., 1989).

2.5 Analytical methods

2.5.1 Cell density and biomass

Optical density was measured with a Cole-Parmer 1100RS spectrophotometer at 600 nm in a cuvette with 10 mm light path. Dry cell weight (dw) was estimated by weighing the samples incubated at 105°C to constant weight.

2.5.2 Determination of substrates and by-products

The concentration of glucose, glycerol, succinate and acetate in the culture medium were determined using a Bruker HPLC system using an Aminex HPX-87H column (Bio-Rad) with both a refractive index (RI) and UV-detector (at 208 nm). Samples were rapidly cooled on ice, then the cells pelleted at 7500g for 5 min and the decanted supernatant was frozen until use. Ten microliters of the supernatant were injected into the column and eluted with 0.009 N H_2SO_4 at a flow rate of 0.5 ml min^{-1} and column temperature of 35ºC. The concentration of ammonium in the culture medium was determined by a colorimetric method using Nessler's reagent (Fluka) as described by Daniels et al. (1994).

2.5.3 Determination of protein concentration

Absorption of motilin, motilin-FP and YUH solutions was determined at 280 nm (A_{280}). Concentration of each protein was calculated using absorption coefficients

deduced from the protein sequence using the software ProtParam available at http://www.expasy.ch/tools/protparam.html (Gasteiger et al., 2005).

2.5.4 Analysis of expression

SDS-PAGE

SDS-PAGE (Laemmli, 1970) was carried out according to Ausubel et al. (1987). 1 ml of OD_{600}=1 culture (or sample with equivalent amount of cells) was centrifuged (2 min, 14000 rpm, RT) and resuspended in 100 µl of loading buffer. Normally, 10 µl of the heated (5 min, 100ºC) sample was loaded per gel slot. Medium molecular mass range markers (Fermentas) were used to localise recombinant proteins.

Immunoblotting

ProtoBlot Western Blot AP system and respective protocol (Promega) were applied to detect his-tag of motilin and YUH. Both, anti-his-tag single chain antibody conjugated to alkaline phosphatase (AP) (obtained from Dr. P. Lindner, University of Zurich Department of Biochemistry) and anti-his-tag rabbit antibody and anti-rabbit antibody AP conjugate were used.

Fluorescence determination

Fluorescence of GST-GFPuv was measured with a Hitachi 840 fluorescence spectrophotometer. Each sample (1 ml of OD=0.05, or containing respective amount of cells) was centrifuged (2 min, 14000 rpm, RT), suspended in 100 mM PBS buffer pH 7.0 containing 0.9% NaCl, and measured. Excitation and emission wavelength maxima were 395 nm and 509 nm, respectively. Fluorescence of GST-GFPuv (in the induced culture, cells or cleared lysate/ pellets during purification) was monitored visually under a UV-lamp, and with a UV-microscope.

2.5.5 Purification procedures

Columns

Separate 5 ml columns were used for purification of YUH, unlabelled and differently labelled motilin-FPs. Columns were packed with Chelating Sepharose Fast Flow (GE Healthcare Life Sciences), loaded with Ni^{2+}, and treated according to the protocols from manufacturer.

Cell lysis by lysozyme treatment and freeze-thaw cycles

Frozen biomass (maintained at -70ºC) was suspended in 4 volumes of lysis buffer A (50 mM Tris, 0.1 M NaCl, pH 8.0), containing lysozyme (Serva) and phenyl methyl sulfonyl fluoride (PMSF) at final concentration of 1 mg ml^{-1} and 1 mM, respectively. Suspension was kept on ice (30 min, 1 h), then three freeze-thaw cycles were carried out: freezing (30 min, -70ºC), thawing at 37ºC with permanent

mixing just until thawed. Alternatively, freeze-thaw cycles were performed by freezing for 1 min in liquid nitrogen, thawing as before. Cell lysis was carried out in 250 ml polypropylene bottles with flat bottoms. To reduce the viscosity, DNase was added to the lysate (at final concentration of 5 μg ml^{-1}) and incubated for 20 min on ice. Viscosity was additionally or alternatively reduced by pumping the lysate through the successively smaller syringe needles. The lysate was centrifuged (27000g, 30 min, 4ºC) and the supernatant (cleared lysate) decanted into a separate tube. The pellets were washed with 3 - 4 volumes (of initial biomass lysed) of buffer A and centrifuged as before. Both supernatants were combined and applied to the column. Washed pellets were maintained at -70ºC until the purification under denaturating conditions.

YUH biomass was lysed as described. β-mercaptoethanol (βME) used in the buffer in French press lysis (VII) was omitted in the current procedure.

Purification of motilin-FP by IMAC in native conditions

Purification was carried out at 0ºC, except centrifugation at 4ºC. The self-packed 5 ml Ni^{2+} NTA-Sepharose column was equilibrated with 50 ml of buffer A, after which cleared cell lysates were applied, and the column was washed with 50 ml of buffer A, followed by another 50 ml of buffer A containing 25 mM of imidazole. The absorbed proteins were eluted with 35 - 50 ml of buffer A containing 250 mM of imidazole. The fractions were kept on ice until enzymatic cleavage with YUH. Each fraction was analysed by SDS-PAGE.

Purification of motilin-FP by IMAC in denaturating conditions

Washed pellets maintained at -70ºC (see above) were suspended in four volumes (of initial biomass lysed) of solubilisation buffer B (buffer A, containing 8 M urea). The suspension was then incubated (1 h, 0ºC) and centrifuged (1 h, 27000g, 4ºC). Pellets were washed by suspension in two volumes (of initial biomass) of buffer B and centrifuged as before. The supernatants from both centrifugations were applied to the column equilibrated with 50 ml of buffer B, column was then washed with another 50 ml of buffer B, followed by 50 ml of buffer A, after with the procedure was carrying on under native conditions.

Preparation of YUH reagent

The IMAC fractions containing YUH were pooled, the working dilution was determined by a 2 - 3 h test-cleavage (using several ratios of motilin-FP and YUH fractions) followed by analysis of the cleavage results by SDS-PAGE. The concentration of YUH solution was determined by absorption at 280 nm. Glycerol and reducing dithiothreitol (DTT) were added to the YUH solution (at final concentration of 40% and 1 mM, respectively), and the resulting reagent was maintained at -20ºC.

Enzymatic cleavage of motilin from motilin-FP by YUH

The fractions containing motilin-FP were pooled, the concentration was determined by measuring absorption at 280 nm, suitable amounts of YUH reagent and DTT (to final concentration of 1 mM) were added and the cleavage solution was incubated at 37ºC for 3 h. The cleavage results were analysed by SDS-PAGE.

Purification of cleavage solution by RP-HPLC

The pH of the cleavage solution was adjusted to 3.5 with 50% TFA (v/ v) and centrifuged (27000g, 4ºC, 20 min). The supernatant was further purified by RP-HPLC and analysed by mass-spectrometry as described in VII. The motilin- (and histidine-tagged ubiquitin) fractions from HPLC were pooled and lyophilised.

3 RESULTS AND DISCUSSION

Optimisation of procedures for the production of labelled motilins are presented and discussed below.

3.1 Expression strategy and production scheme

The expression strategy chosen for the production of labelled motilin used two expression vectors. One vector was used for the production of a decahistidine-tagged ubiquitin-motilin fusion protein (motilin-FP) and the other for the preparation of yeast ubiquitin hydrolase (YUH) required for enzymatic cleavage of motilin from motilin-FP that is not commercially available. Each of the vectors (Fig. 1) was transformed into host strain *E. coli* BL21(DE3). Thus, two recombinant strains were obtained with the expression of target proteins verified using SDS-PAGE and Western blotting (Fig. 2). The motilin-FP was recognized using anti-his-tag rabbit antibodies and anti-rabbit antibodies AP-conjugate as primary and secondary antibodies but not by direct detection using anti-his-tag single chain antibodies conjugated to alkaline phosphatase. The basal expression level varied depending on the cultivation conditions and host from 10 to 100% in both constructs (see below).

The production scheme for labelled motilins shown in Fig. 3 was chosen based on literature analysis and experimental studies (Kohno et al, 1998; VII, this thesis). Development and optimisation of this multistep procedure for the production of ^{15}N-, $^{13}C^{15}N$- and $^{2}H^{15}N$-motilins is described below.

3.2 Cultivation and expression conditions

3.2.1 Choosing the expression host

Good expression levels of motilin-FP and YUH were obtained in each *E. coli* host BL21(DE3) and BL21(DE3) pLysS. BL21(DE3) was chosen for production despite higher basal expression level to avoid an additional antibiotic to maintain the pLysS plasmid and possible complications in fermentation (decrease of growth rate

and expression level (Studier, 1991; Terpe, 2006)) especially using fully deuterated medium.

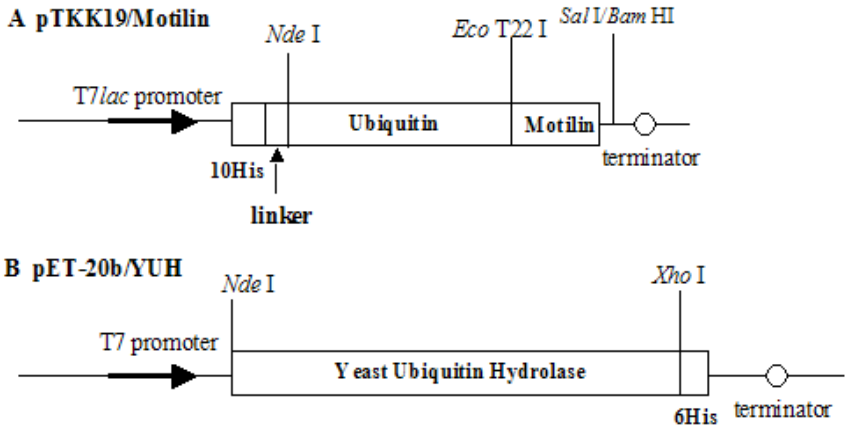

Fig. 1. Expression vectors for decahistidine-tagged ubiquitin-motilin fusion protein (A) and hexahistidine-tagged yeast ubiquitin hydrolase (B) (Kohno et al., 1998; VII).

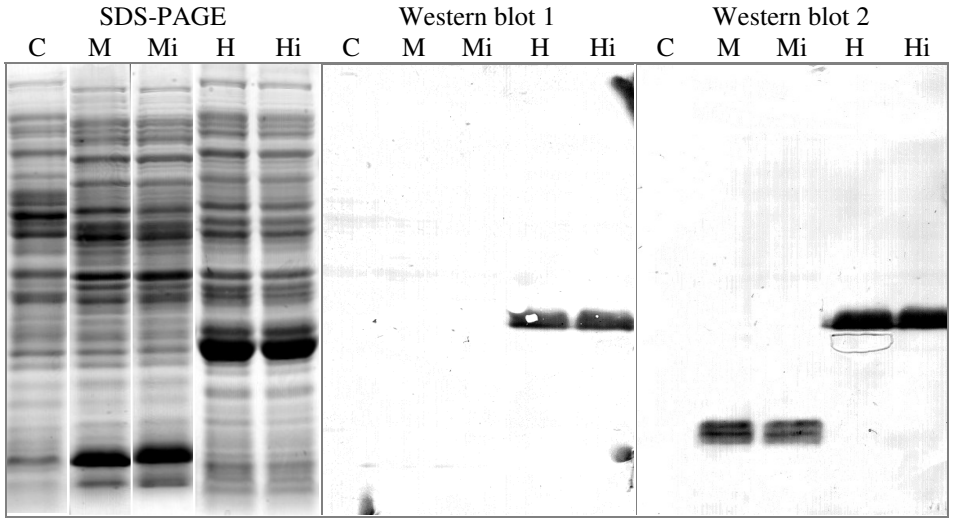

Fig. 2. Verification of motilin-FP and YUH by SDS-PAGE and Western blotting. In Western blot: 1 - anti-his-tag single chain antibodies conjugated to alkaline phosphatase (AP) and 2 - anti-his-tag rabbit antibodies + anti-rabbit antibodies AP-conjugates were used, respectively. C - BL21(DE3); M, Mi and H, Hi - uninduced and induced cultures expressing motilin-FP (13.9 kDa) and YUH (27.2 kDa), respectively.

In numerous cases we noticed that basal expression levels can vary between cultures started from different single colonies but also in the same culture during consecutive cultivation stages (media and growth condition changes). In each case

the basal expression level (~10-100% from final level) was substantial in both motilin-FP and YUH constructs despite the use of T7*lac* promoter in the former. The high basal expression level was common for most target proteins expressed by us (Table 1). The only construct with no basal expression was the GST-GFPuv in pET42a(+) vector under T7*lac* promoter control that was tested using a highly sensitive fluorescence measurement.

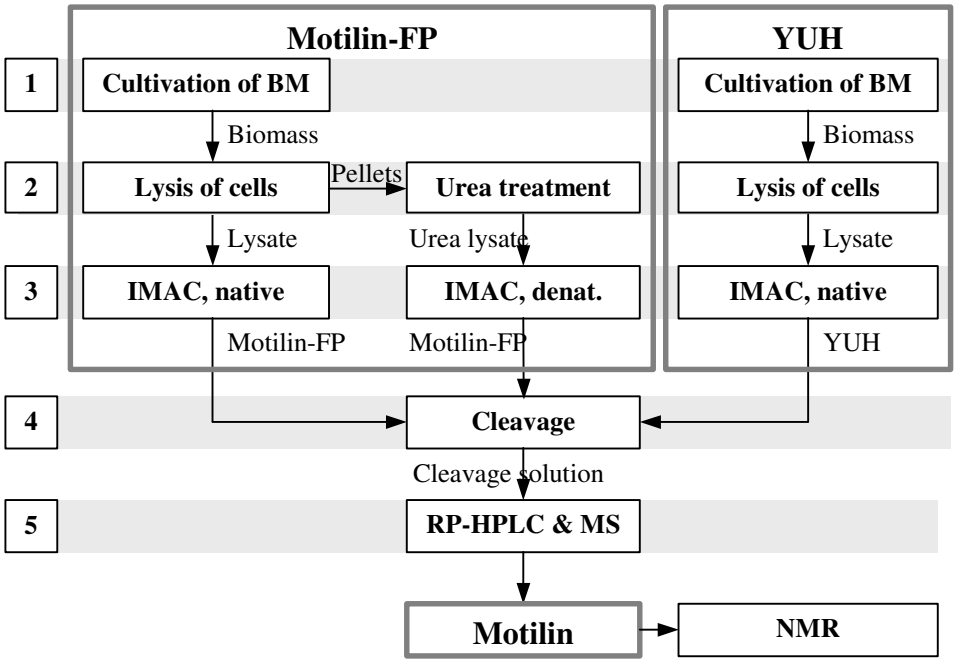

Fig. 3. Scheme for production of (labelled) motilins

3.2.2 Effect of temperature on growth of host and recombinant cultures and on expression of motilin-FP

The effect of temperature was studied by growing the host BL21(DE3) (as control) and the recombinant BL21(DE3) pMot cultures in M9 medium supplemented with succinic (2 g l^{-1}) and acetic (0.5 g l^{-1}) acids. Maximal specific growth rates were similar for both the host (0.07, 0.18, 0.25 h^{-1}) and recombinant culture (0.07, 0.19, 0.21 h^{-1}) grown at 20, 30 and 37ºC respectively.

Optimal results in terms of growth rate, expression rate and level as well as basal expression were obtained at 30ºC (Fig. 4), which was then chosen as the growth and expression temperature for further experiments. Similar results were obtained for other recombinants listed in Table 1. Good expression was also obtained at 20°C, but growth and expression rate were about two times lower.

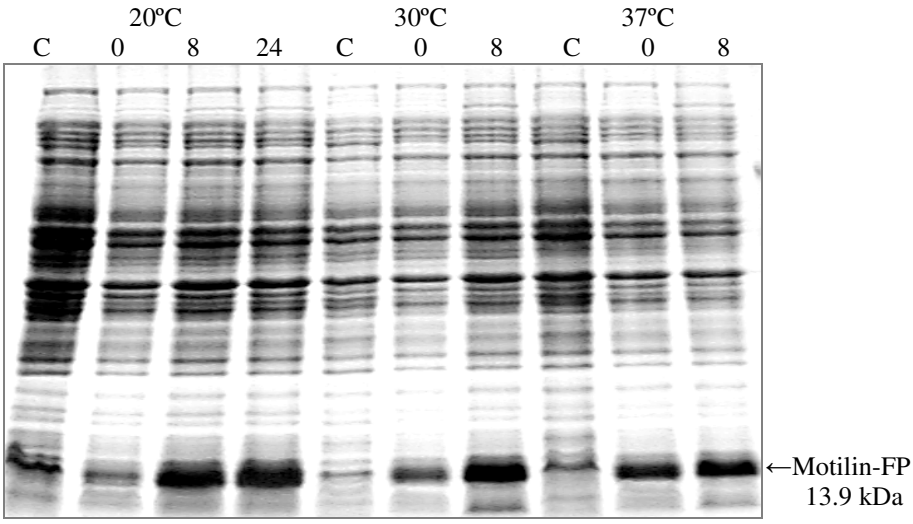

Fig. 4. Motilin-FP expression at different temperatures. SDS-PAGE analysis of total cell lysate: C - BL21(DE3); 0, 8 and 24 - BL21(DE3)pMot, induced with IPTG for 0, 8 and 24 h, respectively.

3.2.3 Selection of carbon source for growth in labelling experiments

The growth of host strain *E. coli* BL21(DE3) on substrates available through chemical synthesis such as glycerol, succinate, succinate+acetate solution, and acetate was studied. The experiments showed that *E. coli* B strains BL21, BL21(DE3) and Origami B (DE3) grow very slowly and have a very long lag phase on succinate compared to *E. coli* strains JM109 and MRE600 (VI). However, on a solution of succinate+acetate the growth rate was higher than on succinate or acetate alone. After acetate was consumed, the growth practically ceased even when succinate was present. Effect of initial acetate concentration on growth is shown in the Fig. 5. Suitable initial concentrations of succinic and acetic acids in batch culture are 1 - 2 g l^{-1} and 0.5 - 1 mg l^{-1}, respectively. Higher acetic acid concentrations were inhibitory.

The poor growth of BL21 using succinate as the sole carbon source can be explained either by the repression of pyruvate kinase or dehydrogenase required for acetyl-CoA synthesis and operation of TCA cycle. In addition, other scientists (Phue and Shiloach 2004; Phue et al., 2005) have shown that the TCA cycle and glyoxylate shunt genes are regulated differently in BL21 and JM109 (*E. coli* B and K strains, respectively). The inhibitory effect of acetate together with poor growth on succinate can explain why the more expensive glucose has been preferred in labelling experiments (Lian and Middleton, 2001; Nietlispach, 2007). On the other hand, we successfully used a high cell density fed-batch/ chemostat method for the production of fully deuterated *E. coli* MRE600 biomass using a solution of

deuterated succinate and acetate (Vanatalu et al., 1993; IV). In the current work the latter method was further developed using an adaptastat technique (VI, and below) and high-density cultures of recombinant BL21(DE3) were obtained on deuterated succinate+acetate mineral media.

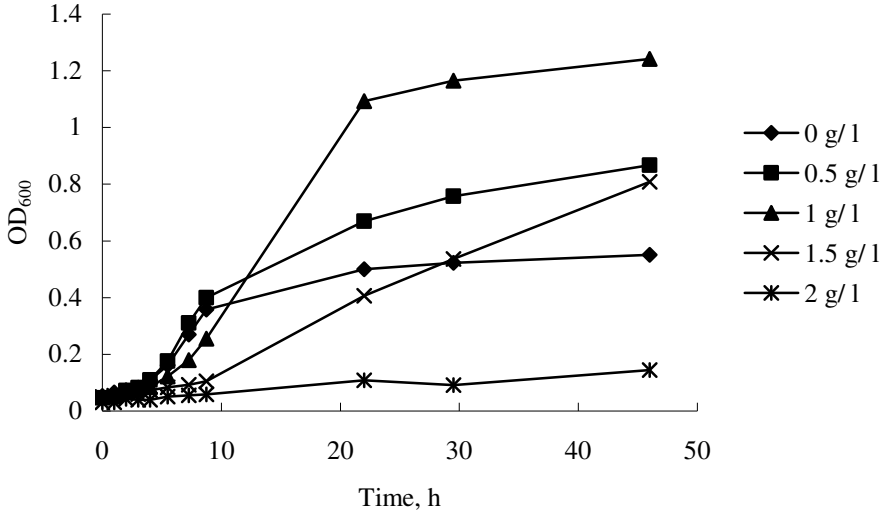

Fig. 5. Effect of initial acetate concentration on growth of *E. coli* BL21(DE3) in batch culture in M9 minimal medium on the solution of succinic and acetic acid (initial concentrations of 2 g l^{-1} and 0 - 2 g l^{-1}, respectively). Optical densities were measured at 600 nm.

3.2.4 Effect of deuteration on growth of recombinant strain BL21(DE3) pMot

The effect of deuteration and different adaptation strategies were studied by growing motilin-FP recombinants in M9 medium containing various percentages of ^2H_2O supplemented with deuterated succinic (2 g l^{-1}) and acetic (0.5 g l^{-1}) acids. Cells grown to the OD~0.4 - 0.5 were inoculated 1: 10 into the next medium. The growth of the cells was severely decreased with increasing percentage of deuteration (Fig. 6). Usually bacteria are adapted to growth in deuterated medium by stepwise subcultures (Moore, 1979; Paliy et al., 2003) where different numbers of steps and ^2H_2O percentages have been used.

In the current study three different adaptation schemes were investigated in order to shorten the protocol from single cell to adapted recombinant fermenter culture (0%→100%; 0%→50%→100%; 0%→70%→100%) (Fig. 6). In the first scheme where the cells (without the medium) were inoculated directly from 0% to 100%-medium the lag phase was longer but the suitable density (~0.4) for the next subcultivation was achieved faster (with less generations) compared with multistep adaptation.

3.2.5 Stability of protein expression in recombinant cultures

Instability is one of the most important and most complicating problems in the cultivation of recombinants and labelling can raise the risk of instability. In paper I, instability was studied theoretically by modelling and using of genetically unstable recombinant producers was discussed. In paper II, instability was studied experimentally during continuous culture of recombinant *E. coli* K12 K802 under unselective conditions. The number of cells resistant toward ampicillin and tetracyclin decreased exponentially (I) from the initial 100% to 3% during the first 48 generations but still about 2% of cells retained resistance during the next 103 generations in continuous culture. According to restriction analysis, the loss of antibiotic resistance marker(s) was not caused by loss of plasmid or by deletions/ insertions in plasmid DNA.

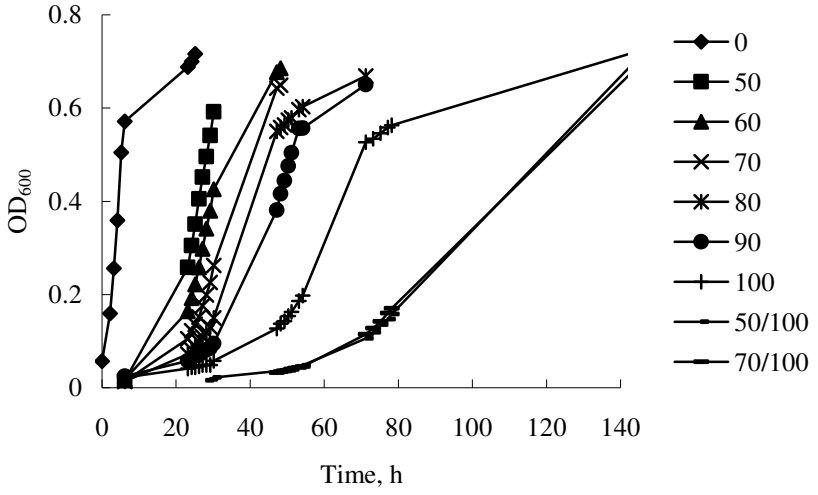

Fig. 6. Effect of deuteration in batch culture of BL21(DE3) pMot in M9 medium containing 2 g l^{-1} of succinic and 0.5 g l^{-1} acetic acids as carbon source and water with different percentage of deuteration. 50/ 100 and 70/ 100 show the experiments where the culture adaptated to 50% or 70% of deuterium was used, respectively.

Like basal expression, instability was encountered in practically all of the recombinants used in this work (Table 1). The loss of recombinant protein synthesis can occur either gradually or quite suddenly during any stage of cultivation. No straightforward solution to this problem exists.

Several practical approaches were used in this work to abate instability. Namely, (1) ensuring of expression stability of a new construct during at least 46 generations which is required to obtain about 20 g dry biomass from a single cell (2) testing the effects of adaptation (to expression conditions on mineral media, deuterated media, substrate limitations, etc) (3) avoiding adverse growth and expression conditions (e.g. stationary phase, starvation, inhibition). The latter is

achieved with proper culture transfer schedules and feeding methods. This is important, because various stresses can increase the mutation rate and affect the expression stability in many ways (Foster, 2005). If the expression was not stable during the required period (generations) the construct was redesigned (human COMT, *E. coli* GroEL&GroES).

3.2.6 Experimental testing of stability

Since it was difficult to obtain good expression levels of recombinant chaperon proteins GroEL and GroES in HCDC in deuterated medium, we studied the feasibility of several constructs/ hosts including constructs with a hok-sok stabilisation mechanism. The testing procedure of *E. coli* MRE600 pPL9amp GroE+hok-sok) is shown in Fig. 7. The cells were adapted to and subsequently grown on fully deuterated medium for 35 generations. The expression level of GroEL was stable but that of GroES decreased between the 33th and 35th generation to about 30% of initial level. With these results, and due to the high variability in expression results among the cultures started from single colonies, the hok-sok constructs were not used for deuteration of GroEL and GroES.

We studied the loss of the cultures ability to express recombinant proteins (e.g. GroEL and GroES) by quantifying and characterising plasmid DNA by isolation, restriction analysis, and agarose electrophoresis. Nevertheless, the loss of expression could not be explained by the data we obtained. This together with the loss of

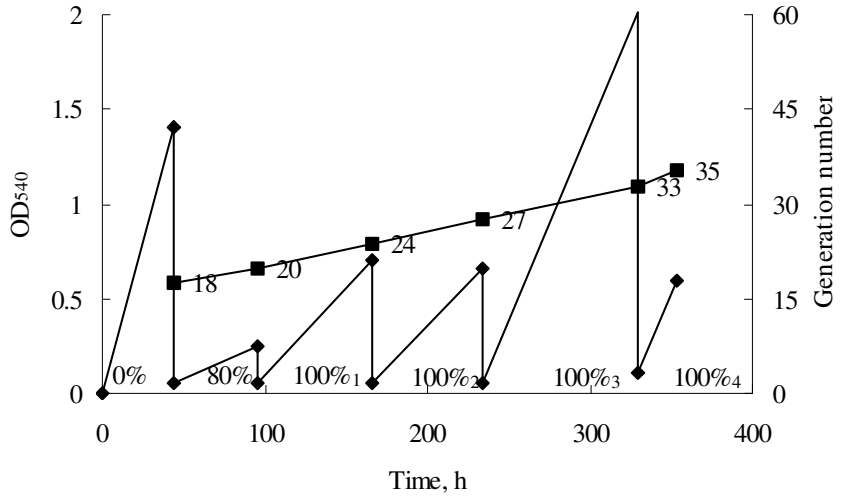

Fig. 7. Cultivation of *E. coli* MRE600 pPL9amp GroE+hok-sok, expressing GroEL and GroES in the expression stability testing experiment. Growth during the adaptation to and subsequent (four) subcultivations on fully deuterated medium. ♦ - OD, measured at 540 nm, ■ - generation number. The cells were grown in 2 ml cultures. The initial OD after inoculation was 0.05.

resistance to antibiotics as in II suggests that the expression instability was not related to loss of plasmids or deletions/ insertions in plasmid DNA.
For practical applications, the method described above was used to assess the expression stability.

3.2.7 Cultivation of active (adapted for growth on deuterated medium) fermentation inoculum

Taking into account the results from studies using different adaptation strategies a scheme was developed for preparation of deuterated inoculum shown in Fig. 8. Several parallel cultures were started from different single cell colonies. Activity was ensured by testing the expression in each subculture in parallel with growth. The final inoculum culture was chosen using expression results from SDS-PAGE. In addition, basal expression level was also considered.

To abate instability, three issues are important. (1) Generation number was minimized during the preparation of inoculation, e.g. by using the shortest adaptation protocol and by transferring as many of cells (including that of single cell colony) as possible into the next step. (2) Stationary growth phase was prevented during inoculum preparation/ adaptation. Physiological state of the cells is important, especially in fully deuterated medium. Expression was often lost or growth was irreversibly stopped if poorly growing (e.g. lag/ stationary phase or inhibited by acetate) cells were used for inoculation. (3) Inoculum size should not be too large or small as either can cause poor growth and even failure of growth, especially in fully deuterated medium. The "safe" size was that giving an initial optical density of 0.05. In addition, brief centrifugation (30 s, 14000 rpm, RT) was used to decrease the amount of H_2O transferred with inoculum into the fermenter for preparation of fully deuterated biomass and recombinant proteins.

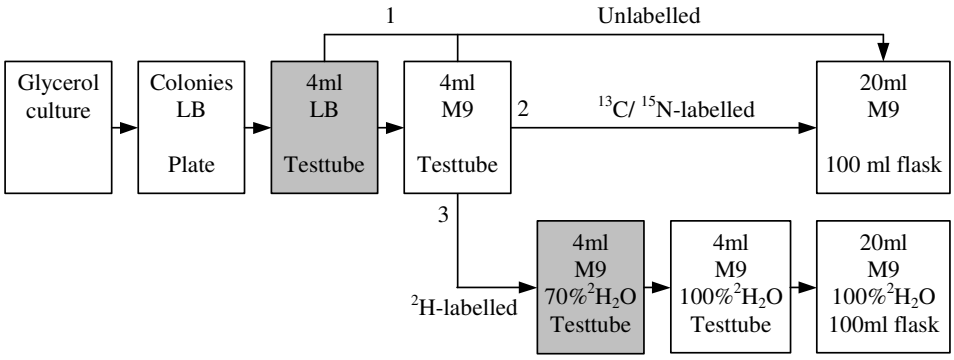

Fig. 8 The scheme of cultivation of active adapted fermentation inoculum. The steps in shadowed boxes were usually skipped in labelling experiments. Three routes for cultivation of inocula are shown.

3.2.8 Conclusions from small-scale batch cultivations

From the experiments, it became obvious that similar expression levels of motilin-FP as well as other recombinant proteins (Table 1) can be achieved independent of media (LB, M9 with various carbon sources glucose, glycerol, succinate+acetate, acetate) and level of deuteration if the expression ability of the culture is not lost during manipulation. This indicates that the main problem of recombinant expression is the instability of the construct.

3.2.9 Optimisation of cultivation and expression conditions in high cell density fed-batch culture

For HCD cultivations batch is unsuitable due to the substrate and by-product inhibition of the process. Substrate inhibition as well as by-product formation due to overflow metabolism can be prevented using carbon source limited fed-batch procedures. For HCD cultivations both appropriate recipes and a fermenter-computer-software system was required (III, V, VII). This fermentation system enabled the use of an appropriate control algorithm that manipulates pH, temperature, dissolved oxygen (DO) and feeding rate.

Feed forward algorithm at predetermined growth rate

In previous HCD expression experiments (III), the μ-stat algorithm (Paalme et al., 1990) was used to obtain high biomass density for recombinant expression. The desired growth rate μ_s was chosen to be equal to about half of the maximum (0.2 h^{-1}) to prevent overflow metabolism and accumulation of inhibitory acetate. In those experiments (III), glycerol was used as a more preferred substrate compared to glucose because glycerol does not cause catabolic repression and overflow metabolism in the case of slight overfeeding.

Effects of several variables (OD at induction, IPTG concentration, postinduction feeding rate, addition of casamino acids to the medium) on recombinant protein expression were investigated in HCD culture (20 g dw l^{-1}). OD at induction and addition of casamino acids had no effect on expression level and rate. It was shown using ^{14}C-labelled glycerol, that recombinant protein was formed from amino acids synthesised *de novo*. At cell densities up to 20 g dw final expression levels similar to that of diluted control was achieved using the same IPTG concentration of 0.5-1 mM for induction of recombinant proteins at both densities. The expression rate was similar at post-induction feeding rates of 0.2 h^{-1} and 0.1 h^{-1} and decreased at 0.05 h^{-1}. The expression level was about 3 times lower in the presence of by-products (at equal feeding rates). The results demonstrated that proper feeding (i.e., avoiding both starvation and by-product formation) before and after induction is another very important factor (besides stability) in achieving good expression levels of recombinant proteins in HCDC.

In pre- and post-induction cultivation phases μ-stat cultivation with predetermined exponential or constant growth rate, respectively was used (III). Under these conditions, high expression levels of recombinant proteins (10-20% of

total cell protein) were obtained. When using HCD we achieved nearly 50-fold higher volumetric yield of several recombinant proteins (Table 1) compared to conventional batch procedures.

Using this method for the expression of deuterated compounds has several restrictions. For producing labelled compounds, the price of glycerol may be too high. Choosing "low price" ^{2}H-labelled substrates like acetic and succinic acid feeding should be very precisely controlled. Low feeding rates result in low growth rate and yield, and high feeding rates result in the accumulation of growth inhibiting acetate, thus decrease the yield. The pre-optimisation of feeding- or μ_S profile is due to the high price of ^{2}H-substrates and absence of micro-scale fed-batch fermenters almost excluded.

To overcome the complication described above and to achieve the "safe" production of labelled recombinant proteins an adaptastat control strategy was developed.

Adaptastat - a new method for real time optimising of growth conditions and cultivation (V)

Our adaptastat method is described in detail in V. The nutritional state (degree of substrate limitation) of culture is tested by periodic lowering of the feeding and by estimating the dissolved oxygen response. According to the response behaviour of DO, the feeding rate was automatically corrected. Using the adaptastat cultivation of *E. coli*, the cells grow at close to maximum growth rate on solutions of succinate+acetate without accumulation of either of substrates in fed-batch culture. The latter is a unique feature, which is very useful in case of labelling, especially deuteration. Optimal feeding can be provided during changing culture conditions of fed-batch culture, e.g. during adaptation procedures and after induction of recombinant protein expression.

One may doubt the effectiveness of this method, especially in the case of recombinant culture, by presuming that periodic interruptions/ lowering in feeding during adaptastat cultivation cause stress to the cells. This was disproved by the fact that the specific growth rate in adaptastat culture was equal to those obtained at low cell densities in batch. In adaptastat, cells can grow with close to the maximal growth rate with no by-product accumulation and there are no differences in culture viability in comparison with continuous culture.

As an example of the feasibility of the method, recombinant *E. coli* BL21(DE3) was cultivated to high density (OD_{600}=24) using an adaptastat fed-batch algorithm in fully deuterated medium on solutions of succinate and acetate. The growth rate of *E. coli* in such difficult conditions was very low (about 0.02 h^{-1}). At higher cell densities, and thus more difficult growth conditions, the specific growth rate started to decrease monotonously. The expression of the reporter protein GST-GFPuv was induced with IPTG at OD_{600}~20. Induction had a clear effect on the utilisation of substrates, as can be seen from a change in feeding profile after the induction. The cells continued to grow with decreasing rate for the next 10 h whereas the GST-GFPuv level increased for 27 h after induction (Fig. 9). A high expression level

was achieved, whereas the specific fluorescence (per OD) of HCD culture was equal to that of the diluted control culture.

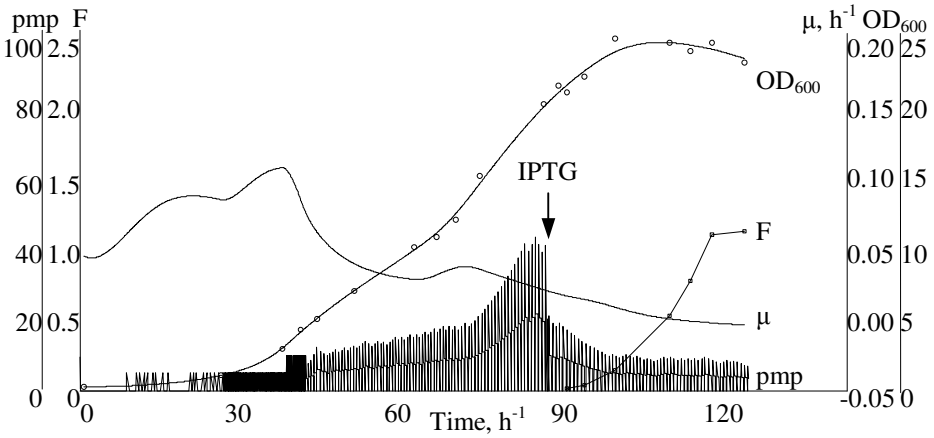

Fig. 9. Fed-batch adaptastat cultivation of *E. coli* BL21(DE3) pET42GFPuv in deuterated medium on the solution of succinic and acetic acids (in acetate limitation). pmp - pumping rate of substrate solution, ml min^{-1}; F - specific fluorescence of GST-GFPuv, relative units; μ - specific growth rate, h^{-1}; OD_{600} - optical density measured at 600 nm.

Production of labelled motilin-FP and YUH biomasses

Our adaptastat method was used for the production of labelled motilin-FPs (VII, hereinafter). Fed-batch fermentation of *E. coli* BL21(DE3) pMOT for the production of $^{2}H^{15}N$-labelled biomass is shown in Fig. 10. This culture was grown in fully deuterated minimal medium with a solution of succinate and acetate, both fully deuterated, used as the carbon and energy source and $^{2}H^{15}N$-labelled ammonium chloride as the nitrogen source. Cells grew with decreasing growth rate (linearly) during the entire adaptastat cultivation. This demonstrates that achieving acceptable growth conditions is a complicated task. Expression of $^{2}H^{15}N$-motilin-FP was induced by adding IPTG on the 95th hour of cultivation. After induction, the cells were cultivated for more than 15 h. During the last two hours the expression level somewhat decreased (Fig. 11). According to data obtained from SDS-PAGE it was decided that the culture should be harvested on the 115th h of cultivation. The feeding was programmed in such a way that all labelled substrates were utilised by the end of cultivation.

Expression of ^{15}N- and $^{13}C^{15}N$-motilin-FP during HCD cultivations analysed by SDS-PAGE is shown in Figs. 12 and 13, respectively.

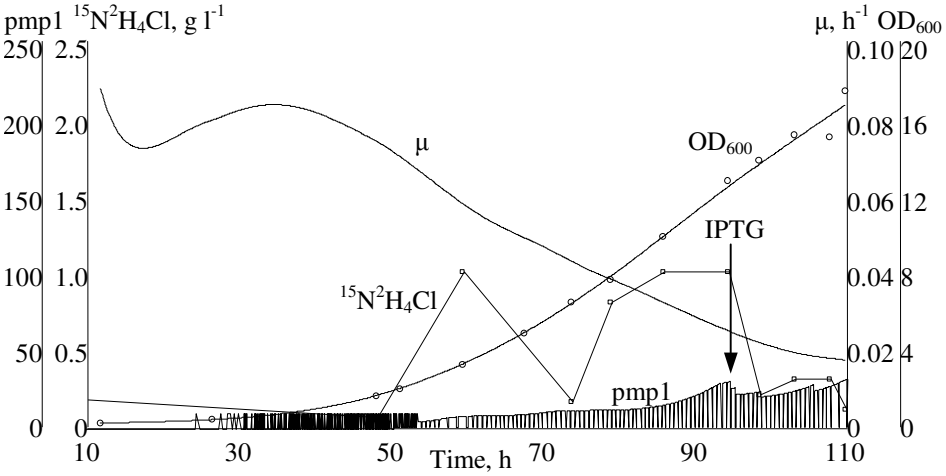

Fig. 10. Fermentation of *E. coli* BL21(DE3) pMOT in fed-batch mode for production of $^2H^{15}N$-labelled biomass. pmp1 - speed of ^2H-succinate+^2H-acetate solution feeding pump (arbitrary units), controlled by adaptastat method; $^{15}N^2H_4Cl$ - concentration of $^{15}N^2H_4Cl$, g l^{-1}; μ - specific growth rate, h^{-1}; OD_{600} - optical density at 600 nm.

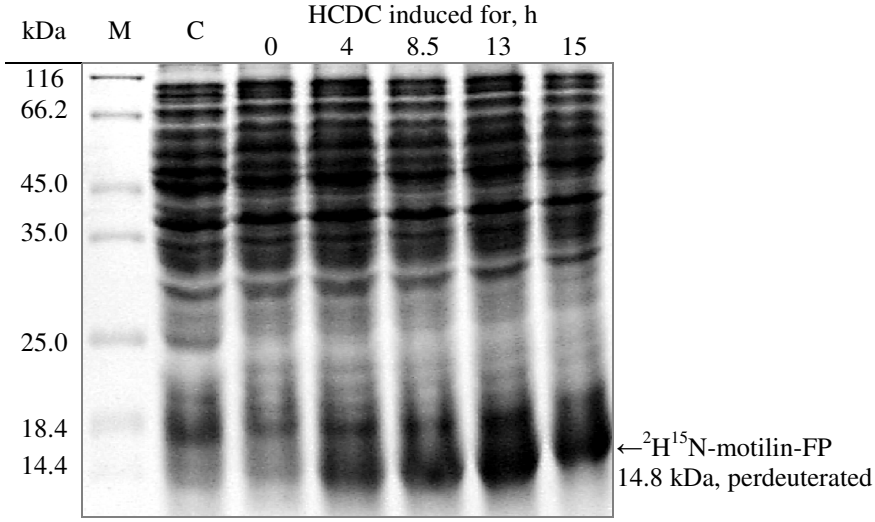

Fig. 11. SDS-PAGE analysis of $^2H^{15}N$-motilin-FP expression in HCD culture of *E. coli* BL21(DE3) pMOT grown in fully deuterated minimal medium on the solution of succinic and acetic acids, and induced with IPTG. M - molecular weight markers; C - control culture BL21(DE3).

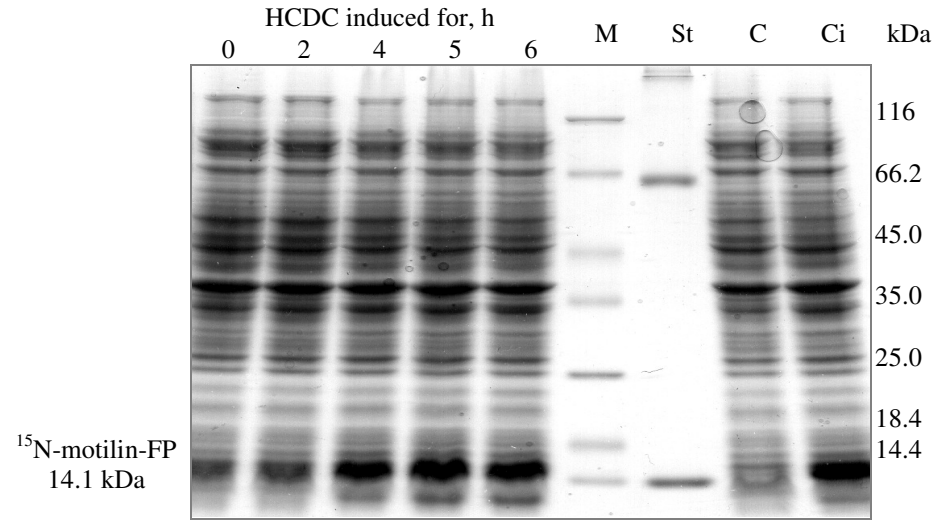

Fig. 12. SDS-PAGE analysis of ^{15}N-motilin-FP expression in HCD culture of *E. coli* BL21(DE3) pMOT grown on glycerol minimal medium, and induced with IPTG ($^{15}NH_4Cl$ was used for labelling). Proteins from 36 µg dw were loaded per well. M - molecular weight markers; St - bovine serum albumin+lysozyme standard, both of 1 µg; C and Ci - uninduced and induced diluted controls, respectively.

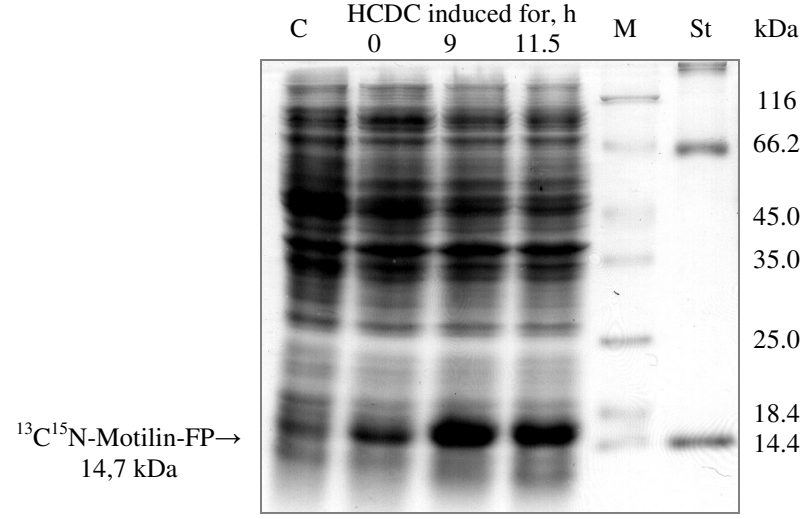

Fig. 13. SDS-PAGE analysis of $^{13}C^{15}$N-motilin-FP expression in HCD culture of *E. coli* BL21(DE3) pMOT grown on ^{13}C-glucose minimal medium, and induced with IPTG ($^{15}NH_4Cl$ was used for ^{15}N-labelling). Proteins from 36 µg dw were loaded per well. C - control culture BL21(DE3); M - molecular weight markers; St - bovine serum albumin+lysozyme standard, both of 1 µg.

3.2.10 Lysis of *E. coli*

French press lysis was used in the purification of unlabelled, ^{15}N- and $^{2}H^{15}N$-motilins, and YUH. An alternative lysis procedure for *E. coli* cell paste (suitable for 5-20 g wet weight) was developed. The procedure for fractionation of cells (Sambrook et al., 1989), consisting of lysozyme treatment and freeze-thaw cycles, was modified with IMAC binding buffer (buffer A) being used instead of original lysis buffer, in addition to different freezing procedures. The feasibility of this procedure was studied by lysis of GST-GFPuv biomass. Lysis efficiency was evaluated by comparing the GST-GFPuv fluorescence in the supernatant and pellets. Optimal lysis buffer to biomass ratio (ml g^{-1} ww) and the number of freeze-thaw cycles were found to be 4-5 and 3, respectively.

Lysis procedure developed was used for preparative lysis of $^{13}C^{15}N$-motilin-FP and YUH biomasses (Table 3).

3.2.11 IMAC purification of his-tagged labelled motilin-FPs and YUH

For the purification of (labelled) motilin-FP and YUH, both histidine-tagged proteins, immobilised Ni^{2+} absorption chromatography was used. Loading, washing and elution conditions, and imidazole concentrations of respective buffers were optimised for motilin-FP and YUH, respectively (VII). In addition, conditions at which stability of motilin-FP and activity of YUH was maintained were estimated (see below).

For comparison of lysis and purification methods, purification yields of unlabelled motilin and $^{13}C^{15}N$-motilin after IMAC (calculated from the absorption measurement data of IMAC eluates) are shown in Table 3. Lysis by freeze-thaw procedure was more efficient with approximately two times higher yield in IMAC under native conditions. Yields of motilin-FP and $^{13}C^{15}N$-motilin-FP were 14.3% and 13.2% of total cell protein, respectively.

Table 3. Comparison of lysis and purification methods

Product	**Lysis and purification conditions, methods**	**Yields after IMAC**	
		mg g^{-1} dw	mg l^{-1}*
Motilin	Native (VII)	6.12	54.2
	Denaturating (VII)	7.44	66.2
$^{13}C^{15}N$-motilin	Native (Materials and Methods)	12.6	90

* Volumetric yield (mg of protein derived per litre of culture volume)

Production of (perdeuterated) $^{2}H^{15}N$-motilin was the most challenging task in this work; therefore, the figures illustrating the production steps as well as NMR-spectra are included as an example.

SDS-PAGE analysis of IMAC purification of $^{2}H^{15}N$-motilin under native conditions is shown in Fig. 14.

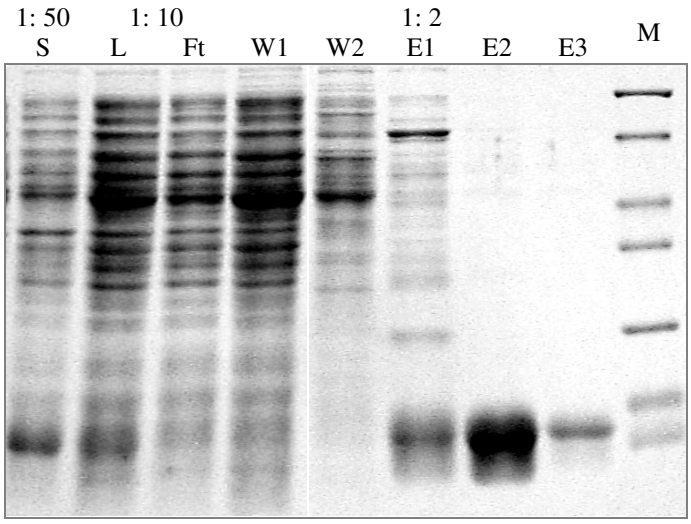

Fig. 14. Purification of histidine-tagged $^{2}H^{15}N$-motilin-FP from 47 g cell paste by immobilized Ni^{2+}absorption chromatography under native conditions, SDS-PAGE analysis. S - cell suspension, L - cleared lysate, Ft - flow through, W1 - 1st wash, W2 - 2nd wash, E1 - E3 - eluted fractions containing $^{2}H^{15}N$-motilin-FP, dilution of samples is shown in the first row. M - molecular weight markers.

3.2.12 Enzymatic cleavage of motilin from motilin-FP by YUH

Stability of motilin-FP and activity of YUH

Both, degradation of motilin-FP and loss of YUH activity were observed in initial experiments. Motilin was lost from the fusion protein either during cell recovery, lysis or IMAC procedures or in one case even during the cultivation. Therefore, we studied fusion protein degradation while adding the serine protease inhibitor PMSF (at final concentration of 0.5 mM), the protease inhibitor cocktail (Roche), and while keeping temperature strictly between 0 - 4ºC. The integrity of motilin-FP was preserved in the cleared (centrifuged) lysates in all of these cases and maintained on ice until purification. However, when the cleared lysates were incubated at 37ºC for 1.5 h, motilin-FP was degraded, but preserved in the similarly treated IMAC eluates. This indicates the degradation of motilin-FP by *E. coli* proteases. Therefore, in addition to the elementary rapid cooling of the culture medium in the end of fermentation, we decided to include PMSF in the lysis buffer in all subsequent experiments to avoid proteolytic degradation, and all the purification steps were carried out at 0ºC (on ice), except centrifugation, which was carried out at 4ºC.

In every motilin-FP purification experiment, YUH was initially purified from the biomass maintained at -70ºC, the IMAC eluate containing

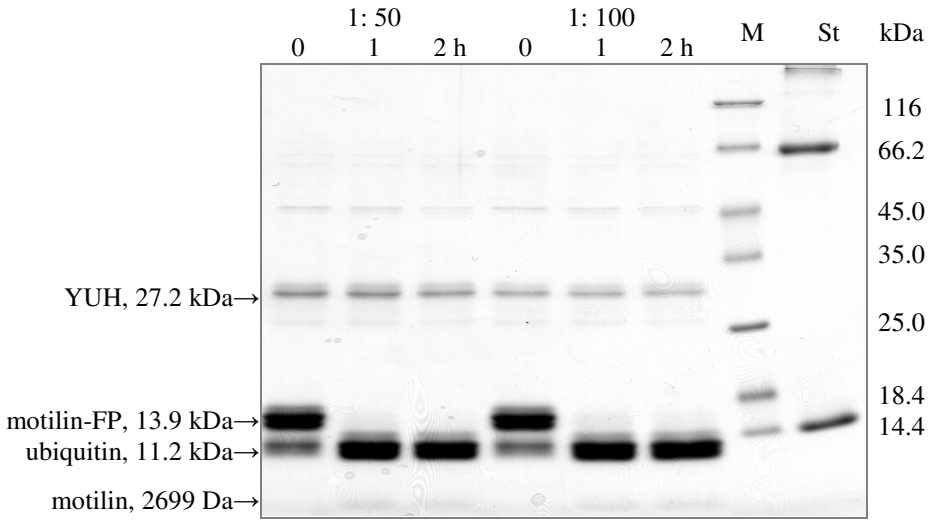

Fig. 15. Cleavage of motilin from motilin-FP by excess of YUH, optimisation of cleavage conditions, SDS-PAGE analysis. Ratios of YUH: motilin-FP solutions in the cleavage solution and incubation times of the latter at 37ºC are shown above the respective lanes. M - Molecular weight markers; St - BSA and lysozyme standards, both 1 µg. YUH, motilin-FP and the fusion partner, i.e. ubiquitin are all his-tagged proteins. Motilin band usually disappeared during destaining see text.

YUH was kept at 4ºC (later at 0ºC) and used for cleavage of motilin-FP in about a week. For each preparative cleavage, both, the ratio of YUH and motilin-FP IMAC eluates, and reaction time was optimised in analytical test cleavage (Fig. 15) on the SDS-PAGE gel. YUH was loosing its activity also at 0ºC during longer periods and successively higher (up to 100 times) concentrations of YUH were needed for the cleavage of motilin from motilin-FP.

Developing and using of YUH reagent for enzymatic cleavage

It was discovered, that the activity of the YUH eluates could be preserved at least for several months by adding 40% glycerol and 1 mM DTT and keeping the solution at -20ºC. Optimal concentrations (mg motilin-FP per mg YUH in 2 h of cleavage) were calculated by using absorption data of both eluates measured at 280 nm. The YUH reagent thus obtained was saved at -20ºC until needed. In addition, it was found to be essential to add 1 mM DTT in the cleavage solution to keep the reduced environment necessary for YUH activity. We determined motilin-FP concentration in the solution of positive IMAC fractions to be cleaved, and the YUH reagent was used at predetermined ratios of proteins. No loss of YUH activity was noticed after 4 months.

Enzymatic cleavage of $^{2}H^{15}N$-motilin from $^{2}H^{15}N$-motilin-FP using YUH reagent is shown in Fig. 16.

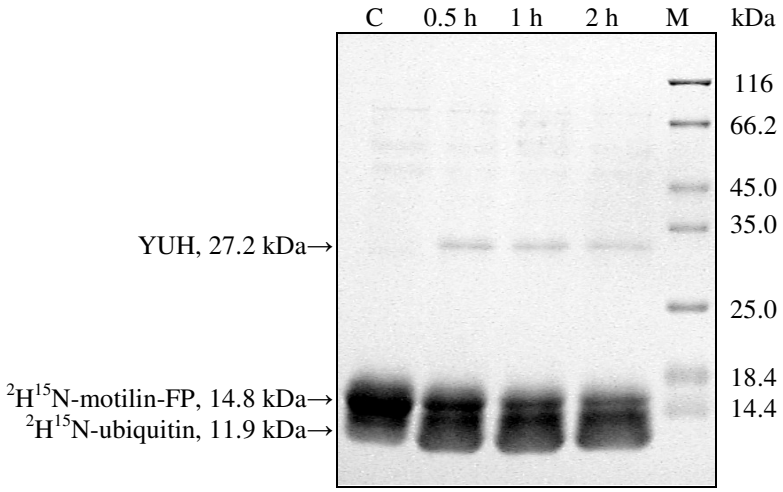

Fig. 16. Enzymatic cleavage of $^{2}H^{15}N$-motilin from $^{2}H^{15}N$-motilin-FP using YUH reagent, SDS-PAGE analysis. C - control (cleavage solution without YUH), on top of the following lanes incubation time of cleavage solution is shown, M - molecular weight markers. Both, perdeuterated $^{2}H^{15}N$-motilin-FP and fusion partner, i.e. $^{2}H^{15}N$-ubiquitin are his-tagged proteins.

3.2.13 Purification of cleavage solution by RP-HPLC, and Mass-Spectrometry

After the cleavage, motilin and his-tagged ubiquitin were purified from the cleavage solution by RP-HPLC (VII). For successful purification by preparative HPLC, it was essential to adjust the pH of the cleaved solution to 3.5 by using TFA. Significant precipitation occurred at this pH, but concentration of higher-weight proteins, excluding motilin, was decreased in the supernatant (estimated by SDS-PAGE and HPLC). The mixture was then centrifuged and the supernatant purified by HPLC.

Purification of $^{2}H^{15}N$-motilin from cleavage solution by RP-HPLC and the deconvulated mass spectrum of purified $^{2}H^{15}N$-motilin (2872 Da) is shown in Figs. 17 and 18, respectively.

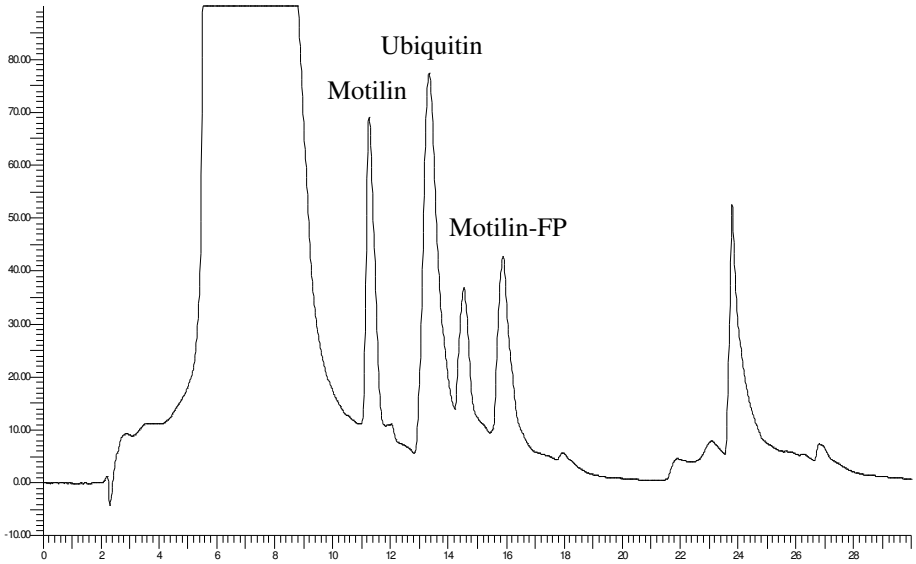

Fig. 17. Purification of $^{2}H^{15}N$-motilin from cleavage solution, RP-HPLC chromatogram.

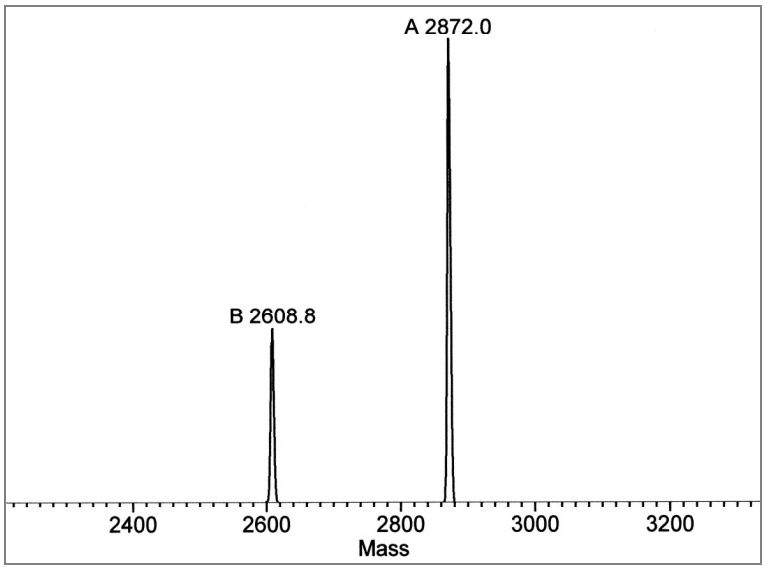

Fig. 18. Deconvulated mass spectrum of purified $^{2}H^{15}N$-motilin, 2872 Da.

3.2.14 Data on labelled motilin production

Summary of production data of products of this study are shown in Tables 3, 4, and 5. Enrichment in (^{13}C, ^{15}N, C-bonded ^{2}H) uniform labelling obtained (Table 5) was practically equal to that of source chemicals (Table 2) and was sufficient for corresponding NMR experiments.

Table 4. Production data of uniformly labelled motilins

Product	Fed-batch fermentation				Purification	
	Carbon/ labelling sources	Induction phase			Final yields	
		OD_{600}*		Hours	mg g^{-1} dw	mg l^{-1}
^{15}N-motilin	Glycerol, $^{15}NH_4Cl$	7.5	13	11	0.52	4.7
^{15}N-motilin	Glycerol, $^{15}NH_4Cl$	7.0	16	6	0.76	10
$^{2}H^{15}$N-motilin	^{2}H-(succinic+acetic) acids, $^{15}N^{2}H_4Cl$, $^{2}H_2O$	13	17	15	0.29	3.9
$^{13}C^{15}$N-motilin	^{13}C-glucose, $^{15}NH_4Cl$	2.5	9.6	11.5	0.76	5.4

* OD_{600} values at the beginning and at the end of induction phase

Table 5. Molecular weights and isotopic enrichments of produced labelled motilins

Product	Molecular weight, Da		Enrichment, % atom of
	Theoretical	Experimental	
Motilin	2699.1	2699.0±0.0	
^{15}N-motilin	2732.8	2732.2±0.4	98.2±1.2
$^{2}H^{15}$N-motilin	2872.7	2872.0±0.3	99.6±0.2
$^{13}C^{15}$N-motilin	2851.5	2849.7±0.3	98.8±0.2

3.2.15 NMR spectra of $^{2}H^{15}$N-motilin

600 MHz ^{1}H NMR spectra of $^{2}H^{15}$N-labeled motilin are shown in Fig. 19. Amide proton exchange is fast and all amide deuterons have been exchanged with protons during the sample preparation. The amide region shows very sharp amide resonances as singlets due to the high degree of deuteration. Without deuteration amide resonances would split into doublets due to 3-bond couplings to α-protons (4 - 9 Hz) (Fig. 19, top). In the spectrum recorded without ^{15}N decoupling (Fig. 19, bottom), all amide resonances show doublets with typical one bond NH coupling (~90 Hz).

600 MHz $^{1}H^{15}$N HSQC spectrum of $^{2}H^{15}$N-labelled motilin shows well-resolved crosspeaks from amino acid residues with only the signals from amino acid residues M13 and E17 overlapping (Fig. 20). Crosspeaks are designated by a one-letter code (for convenience, motilin sequence in three-letter code: Phe-Val-Pro-Ile-Phe-Thr-Tyr-Gly-Glu-Leu-Gln-Arg-Met-Gln-Glu-Lys-Glu-Arg-Asn-Lys-Gly-Gln). N-terminal phenylalanine and proline do not contribute to the spectrum. Weak signals next to G21 and Q22 are due to impurity.

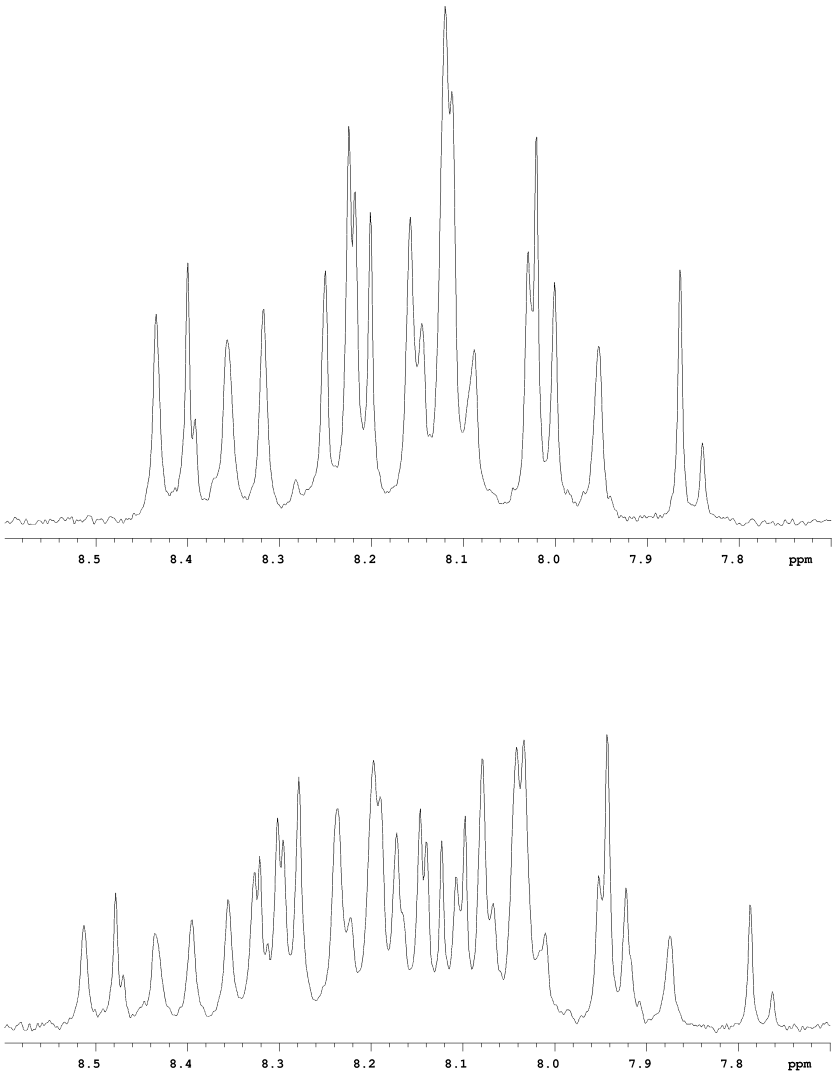

Fig. 19. 600 MHz ^{1}H NMR spectra of $^{2}H^{15}N$-labeled motilin in 90% H_2O/ 10% $^{2}H_2O$ at 25°C, pH 3.6. Top: amide region ^{15}N decoupled, bottom: without ^{15}N decoupling.

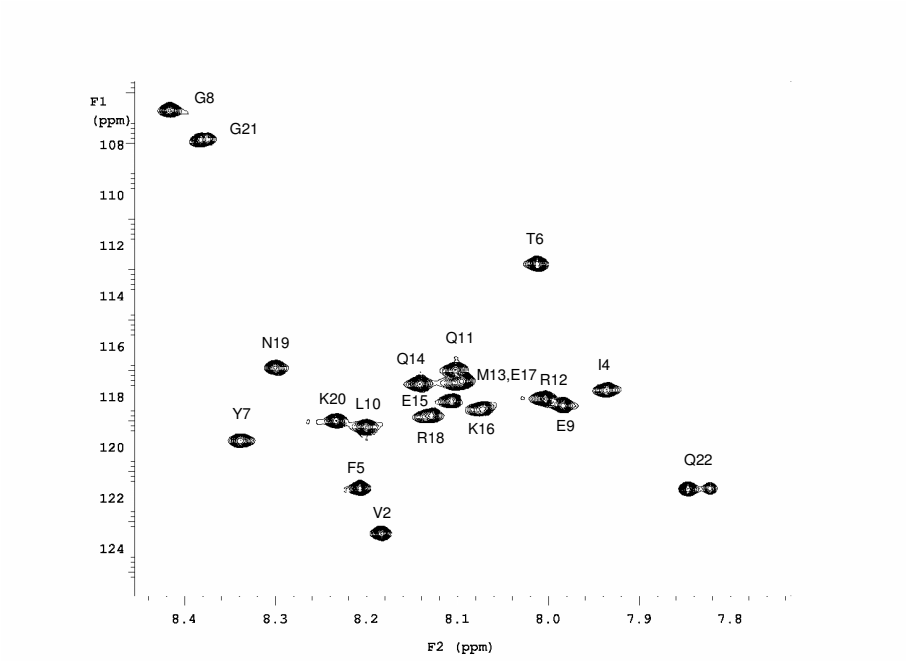

Fig. 20. 600 MHz $^{1}H^{15}N$ HSQC spectrum of $^{2}H^{15}N$-labelled motilin in 90% H_2O/ 10% $^{2}H_2O$ at 25°C, pH 3.6. Amino acid residue crosspeaks are labelled using one-letter code.

CONCLUSIONS

1. The automatic fed-batch procedures developed enabled us to obtain high expression yields of recombinant proteins in *E. coli* at high cell densities and deuteration levels.
2. Glycerol is a convenient and "safe" carbon source for expression of recombinant proteins in *E. coli* that enables high specific expression levels (10-20% of total cell protein) at high cell densities (20 g dw l^{-1}) (Table 1), and increases the volumetric yield compared to the conventional (batch) procedures by nearly 50-fold (III).
3. A new method for cultivation of microbes at maximum growth rate adaptastat (V) solves the problem of controlling the feeding rate at an optimal level in case of isotopic, particularly ^{2}H-labelling.
4. It was discovered, that common expression hosts *E. coli* BL21(DE3) and *E. coli* Origami B (DE3) differ from *E. coli* strains JM109 and MRE600, by growing poorly on succinate as a sole carbon source, however they grew reasonably well on the acetate+succinate solution. This allows using these simple synthetic compounds as cost-effective substrates for the production of labelled recombinant proteins (VI).

5. The final expression level (mg g dw^{-1}) of recombinant proteins is generally only slightly affected by factors that decrease the growth rate - low temperature and feeding rate, nonglycolytic substrates, high degree of deuteration, and high biomass densities.
6. Using these methods, the unlabelled and labelled biomass of *E. coli* strains expressing a number of recombinant proteins (listed in Table 1) was produced in 10-100 g ww amounts. The costs for expression and labelling were minimised (e.g. by using less expensive substrates and saving heavy water in HCD/ low volume deuterated culture). Most of the recombinant strains and constructs were provided by our collaboration partners and the respective unlabelled or labelled biomasses produced were given back to our partners for subsequent purification and protein structure studies.
7. ^{15}N, $^{13}C^{15}N$ and $^{2}H^{15}N$ isotopic *in vivo* labelling of the hormone peptide motilin in fed-batch cultivation were developed. In addition, a method for purifying the labelled recombinant proteins was developed, and consisted of cell lysis, affinity chromatography and proteolytic cleavage steps. Preparative amounts of ^{15}N-, $^{13}C^{15}N$- and $^{2}H^{15}N$-labelled motilins were purified. (Labelled motilins and ubiquitins were further purified by RP HPLC, and from the purified peptides with different isotopic labelling heteronuclear NMR spectra were obtained (all done by collaboration partners) which enabled them to study the structure and dynamics of motilin. The larger aim of the motilin project was to determine the receptor bound conformation of peptide ligands. This work is underway and more labelled motilin is required.)

REFERENCES

Actis LA, Tolmasky ME and Crosa JH (1999) Bacterial plasmids: replication of extrachromosomal genetic elements encoding resistance to antimicrobial compounds. *Frontiers in Bioscience* 3, d43-62

Åkesson M, Hagander P, Axelsson JP (2001) Avoiding acetate accumulation in *Escherichia coli* cultures using feedback control of glucose feeding. *Biotechnology and Bioengineering* 73, 223-230

Altenbuchner J and Mattes R (2005) *Escherichia coli*. In: G.Gellissen (ed.) Production of Recombinant Proteins, Wiley-VCH, 7-43

Atlung T, Nielsen A, Rasmussen LJ, Nellemann LJ, Holm F (1991) A versatile method for integration of genes and gene fusions into the lambda attachment site of *Escherichia coli*. *Gene* 107, 11-17

Ausubel FM, Brent R, Kingston RE et al. (1987) Current Protocols in Molecular Biology, John Wiley, NY

Baker KE, Mackie GA (2003) Ectopic RNase E sites promote bypass of 5'-end-dependent mRNA decay in *Escherichia coli*. *Molecular Microbiology* 47, 75-88

Baneyx F, Palumbo JL (2003) Improving heterologous protein folding via molecular chaperone and foldase co-expression. *Methods in Molecular Biology* 205, 171-197

Baneyx F (1999) Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology* 10, 411-421

Baneyx F, Mujacic M (2004) Recombinant protein folding and misfolding in *Escherichia coli*. *Nature Biotechnology* 22, 1399-1408

Ben-Bassat A (1991) Methods for removing N-terminal methionine from recombinant proteins. *Bioprocess Technology* 12, 147-159

Bentley WE, Mirjalili N, Andersen DC, Davis RH and Kompala DS (1990) Plasmid-encoded protein: The principal factor in the metabolic burden associated with recombinant bacteria. *Biotechnology and Bioengineering* 35, 668-681

Berman H, Henrick K, Nakamura H (2003) Announcing the worldwide Protein Data Bank. Nature Structural Biology 10, 980

Bessette PH, Aslund F, Beckwith J, Georgiou G (1999) Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proceedings of the National Academy of Sciences USA* 96, 13703-13708

Betenbaugh MJ, Beaty C, Dhurjati P (1989) Effects of plasmid amplification and recombinant gene expression on the growth kinetics of recombinant *E. coli*. *Biotechnology and Bioengineering* 33, 1425-1436

Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453-1474

Boe L, Tolker-Nielsen T (1997) Plasmid stability: comments on the dimer catastrophe hypothesis. *Molecular Microbiology* 23, 247-253

Bornhorst JA and Falke JJ (2000) Purification of proteins using polyhistidine affinity tags. *Methods in Enzymology* 326, 245-254

Cai M, Huang Y, Sakaguchi K, Clore GM, Gronenborn AM, Craigie R (1998) An efficient and cost-effective isotope labeling protocol for proteins expressed in *Escherichia coli*. *Journal of Biomolecular NMR* 11, 97-102

Calos MP (1978) DNA sequence for a low-level promoter of the lac repressor gene and an 'up' promoter mutation. *Nature* 274, 762-765

Carrió MM, Corchero JL, Villaverde A (1998) Dynamics of in vivo protein aggregation: building inclusion bodies in recombinant bacteria. *FEMS Microbiology Letters* 169, 9-15

Carrió MM, Villaverde A (2001) Protein aggregation as bacterial inclusion bodies is reversible. *FEBS Letters* 489, 29-33

Carver J, Keniry MA (2002) NMR spectroscopy of large proteins. *Annual Reports on NMR Spectroscopy* 48, 32-69

Chant EL and Summers DK (2007) Indole signalling contributes to the stable maintenance of *Escherichia coli* multicopy plasmids. *Molecular Microbiology* 63, 35-43

Choi JH, Keum KC, Lee SY (2006) Production of recombinant proteins by high cell density culture of *Escherichia coli*. *Chemical Engineering Science* 61, 876-885

Choi JH Lee SY (2004) Secretory and extracellular production of recombinant proteins using *Escherichia coli*. *Applied Microbiology and Biotechnology* 64, 625-635

Choy JS, Aung LL, Karzai AW (2007) Lon protease degrades transfer-messenger RNA-tagged proteins. *Journal of Bacteriology* 189, 6564-6571

Chung BH, Choi YJ, Yoon SH, Lee SY and Lee YI (2000) Process development for production of recombinant human insulin-like growth factor-I in *Escherichia coli*. *Journal of Industrial Microbiology and Biotechnology* 24, 94-99

Cohen SN, Chang ACY, Boyer HW and Helling RB (1973) Construction of biologically functional bacterial plasmids in vitro. *Proceedings of The National Academy of Sciences USA* 70, 3240-3244

Corchero JL, Villaverde A (1998) Plasmid maintenance in *Escherichia coli* recombinant cultures is dramatically, steadily, and specifically influenced by features of the encoded proteins. *Biotechnology and Bioengineering* 58, 625-632

Crameri A, Whitehorn EA, Tate E and Stemmer WPC (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature Biotechnology* 14, 315-319

Daniels L, Hanson RS and Phillips JA (1994) Chemical analysis. In: Methods for General and Molecular Bacteriology (Gerhardt P, Murray RGE, Wood WA and Krieg NR eds) Washington, DC: American Society for Microbiology, ch 22, 512-554

Datar RV, Cartwright T and Rosen C-G (1993) Process economics of animal cell and bacterial fermentations: a case study analysis of tissue plasminogen activator. Bio/Technology 11, 349-357

Davies SJ, Golby P, Omrani D, Broad SA, Harrington VL, Guest JR, Kelly DJ, and Andrews SC (1999) Inactivation and Regulation of the Aerobic C4-Dicarboxylate Transport (dctA) Gene of *Escherichia coli*. *Journal of Bacteriology* 181, 5624-5635

Desplancq D, Bernard C, Sibler AP, Kieffer B, Miguet L, Potier N, Van Dorsselaer A, Weiss E (2005) Combining inducible protein overexpression with NMR-grade triple isotope labeling in the cyanobacterium *Anabaena* sp. PCC 7120. *Biotechniques* 39, 405-411

Deuerling E, Patzelt H, Vorderwulbecke S, Rauch T, Kramer G, Schaffitzel E, Mogk A, Schulze-Specking A, Langen H, Bukau B (2003) Trigger Factor and DnaK possess overlapping substrate pools and binding specificities. *Molecular Microbiology* 47, 1317-1328

Dong H, Nilsson L, Kurland CG (1996) Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *Journal of Molecular Biology* 260, 649-663

Dubendorff JW and Studier FW (1991) Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *Journal of Molecular Biology* 219, 45-59

Dyck MK, Lacroix D, Pothier F and Sirard M-A (2003) Making recombinant proteins in animals – different systems, different applications. *Trends in Biotechnology* 21, 394-9

Ebersbach G and Gerdes K (2005) Plasmid Segregation Mechanisms. *Annual Review of Genetics* 39, 453-479

Eiteman MA and Altman E (2006) Overcoming acetate in *Escherichia coli* recombinant protein fermentations. *Trends in Biotechnology* 24, 530-536

Ellis RJ and Minton AP (2003) Join the crowd. *Nature* 425, 27-28

ESF Report (2001) Protein Structure and Function in the Post Genomic Era. European Scientific Foundation Study Report, www.esf.org

Ewalt KL, Hendrick JP, Houry WA, Hartl FU (1997) In vivo observation of polypeptide flux through the bacterial chaperonin system. *Cell* 90, 491-500

Fahnert B, Lilie H and Neubauer P (2004) Inclusion bodies: formation and utilisation. *Advances in Biochemical Engineering/Biotechnology* 89, 93-142

Farewell A and Neidhardt FC (1998) Effect of temperature on in vivo protein synthetic capacity in *Escherichia coli*. *Journal of Bacteriology* 180, 4704-4710

Fernández LA, Sola I, Enjuanes L, de Lorenzo V (2000) Specific secretion of active single-chain Fv antibodies into the supernatants of *Escherichia coli* cultures by use of the hemolysin system. *Applied and Environmental Microbiology* 66, 5024-5029

Fiedler S, Knocke C, Vogt J, Diehl A (2007) Production of ^{2}H-, ^{13}C-, and ^{15}N-labeled OmpG via high cell density fermentation. *Genetic Engineering & Biotechnology News* 27(9)

Fiedler M, Skerra A (2001) proBA complementation of an auxotrophic *E. coli* strain improves plasmid stability and expression yield during fermenter production of a recombinant antibody fragment. *Gene* 274, 111-118

Foster PL (2005) Stress responses and genetic variation in bacteria. *Mutation Research* 569, 3-11

Friehs K (2004) Plasmid copy number and plasmid stability. *Advances in Biochemical Engineering/Biotechnology* 86, 47-82

Garcia-Fruitos E, Gonzalez-Montalban N, Morell M, Vera A, Ferraz RM, Aris A, Ventura S and Villaverde A (2005) Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microbial Cell Factories* 4: 27

Gardner KH and Kay LE (1998) The use of ^{2}H, ^{13}C, ^{15}N multidimensional NMR to study the structure and dynamics of proteins. *Annual Review of Biophysics and Biomolecular Structure* 27, 357-406

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. In: Walker JM (ed) The Proteomics Protocols Handbook, Humana Press, 571-607

Gellissen G, Strasser AWM, Suckow M (2005) Key and criteria to the selection of an expression platform. In: Gellissen G (ed) Production of recombinant proteins – Novel microbial and eukaryotic expression systems, Weinheim: Wiley-VCH, 1-5

Georgiou G and Segatori L (2005) Preparative expression of secreted proteins in bacteria: status report and future prospects. *Current Opinion in Biotechnology* 16, 538-545

Gerdes, K (1988) The parB (hok/sok) locus of plasmid R1: a general purpose plasmid stabilization system. *Bio/Technology* 6, 1402-1405

Glascock CB, Weickert MJ (1998) Using chromosomal lacIQ1 to control expression of genes on high-copy-number plasmids in *Escherichia coli*. *Gene* 223, 221-231

Glick BR (1995) Metabolic load and heterologous gene expression. *Biotechnology Advances* 13, 247-261

Goeddel DV, Kleid DG, Bolivar F, Heyneker HL, Yansura DG, Crea R, Hirose T, Kraszewski A, Itakura K, and Riggs AD (1979) Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proceedings of The National Academy of Sciences USA* 76, 106-110

Goto NK, Kay LE (2000) New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Current Opinion in Structural Biology* 10, 585-592

Grabherr R, Nilsson E, Striedner G, Bayer K (2002) Stabilizing plasmid copy number to improve recombinant protein production. *Biotechnology and Bioengineering* 77, 142-147

Grossman TH, Kawasaki ES, Punreddy SR, Osburne MS (1998) Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. *Gene* 209, 95-103

Han K, Hong J, Lim HC (1993) Relieving effects of glycine and methionine from acetic acid inhibition in *Escherichia coli* fermentations. *Biotechnology and Bioengineering* 41, 316-324

Hannig G, and Makrides SC (1998) Strategies for optimizing heterologous protein expression in *Escherichia coli*. *Trends in Biotechnology* 16, 54-60

Haon S, Augé S, Tropis M, Milon A and Lindley N (1993) Low cost production of perdeuterated biomass using methylotrophic yeasts. *Journal of labelled compounds and radiopharmaceuticals* 33, 1053-1063

Harcum SW (2002) Structured model to predict intracellular amino acid shortages during recombinant protein overexpression. *Journal of Biotechnology* 93, 189-202

Harcum SW, Bentley WE (1999) Heat-shock and stringent responses have overlapping protease activity in *Escherichia coli*. Implications for heterologous protein yield. *Applied Biochemistry and Biotechnology* 80, 23-37

Held D, Yaeger K and Novy R (2003) New coexpression vectors for expanded compatibilities in *E. coli*. *InNovations* 18, 4-6

Hellmuth K, Korz DJ, Sanders EA, Deckwer W-D (1994) Effect of growth rate on stability and gene expression of recombinant plasmids during continuous and high cell density cultivation of *Escherichia coli* TG1. Journal of Biotechnology 32, 289-298

Hellwig S, Drossard J, Twyman RM & Fischer R (2004) Plant cell cultures for the production of recombinant proteins. *Nature Biotechnology* 23, 1415-1422

Herendeen SL, VanBogelen RA, Neidhardt FC (1979) Levels of major proteins of *Escherichia coli* during growth at different temperatures. *Journal of Bacteriology* 139, 185-194

Hoffmann F and Rinas U (2004) Stress induced by recombinant protein production in *Escherichia coli*. *Advances in Biochemical Engineering/Biotechnology* 89, 73-92

Holms H (1996) Flux analysis and control of the central metabolic pathways in *Escherichia coli*. *FEMS Microbiology Reviews* 19, 85-116

Holzinger J, Heumann H, Manakova E, Rössle M, K Vanatalu K, Wiedenmann A, May RP (2000) GroEL single-ring solution structure obtained by small-angle neutron scattering. *Physica B* 276-278, 528-529

Horn U, Strittmatter W, Krebber A, Knüpfer U, Kujau M, Wenderoth R, Müller K, Matzku S, Plückthun A, Riesenberg D (1996) High volumetric yields of functional dimeric miniantibodies in *Escherichia coli*, using an optimized expression vector and high-cell-density fermentation under non-limited growth conditions. *Applied Microbiology and Biotechnology* 46, 524-532

Hua Q, Yang C, Oshima T, Mori H, Shimizu K (2004) Analysis of gene expression in *Escherichia coli* in response to changes of growth-limiting nutrient in chemostat cultures. *Applied and Environmental Microbiology* 70, 2354-2366

Ilgen C, Lin-Cereghino J, Cregg JM (2005) Pichia pastoris. In: Gellissen G (ed) Production of recombinant proteins – Novel microbial and eukaryotic expression systems. Weinheim: Wiley-VCH, 143-162

Inouye M (2006) The discovery of mRNA interferases: implication in bacterial physiology and application to biotechnology. *Journal of Cellular Physiology* 209, 670-676

Inouye M, Arnheim N, Sternglanz R (1973) Bacteriophage T7 lysozyme is an N-acetylmuramyl-L-alanine amidase. *The Journal of Biological Chemistry* 248, 7247-7252

Itakura K, Hirose T, Crea R, Riggs AD, Heyneker HL, Bolivar F, Boyer HW (1977) Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science* 198, 1056-1063

Jackson DA, Symons RH, Berg P (1972) Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proceedings of The National Academy of Sciences USA* 69, 2904-2909

Jonasson P, Liljeqvist S, Nygren P-Å and Ståhl S (2002) Genetic design for facilitated production and recovery of recombinant proteins in *Escherichia coli*. *Biotechnology and Applied Biochemistry* 35, 91-105

Kane JF (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Current Opinion in Biotechnology* 6, 494-500

Kohno T, Kusunoki H, Sato K, and Wakamatsu K (1998) A new general method for the biosynthesis of stable isotope-enriched peptides using a decahistidine-tagged ubiquitin fusion system: An application to the production of mastoparan-X uniformly enriched with ^{15}N and $^{15}N/^{13}C$. *Journal of Biomolecular NMR* 12, 109-121

Korpimäki T, Kurittu J, Karp M (2003) Surprisingly fast disappearance of β-lactam selection pressure in cultivation as detected with novel biosensing approaches. *Journal of Microbiological Methods* 53, 37-42

Korz DJ, Rinas U, Hellmuth K, Sanders EA, Deckwer WD (1995) Simple fed-batch technique for high cell density cultivation of *Escherichia coli*. *Journal of Biotechnology* 39, 59-65

Kroef EP, Owens RA, Campbell EL, Johnson RD and Marks HI (1989) Production scale purification of biosynthetic human insulin by reversed-phase high-performance liquid chromatography. *Journal of Chromatography* 461, 45-61

Kurland C, Gallant J (1996) Errors of heterologous protein expression. *Current Opinion in Biotechnology* 7, 489-493

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

Lee J, Lee SY, Park S, Middelberg APJ (1999) Control of fed-batch fermentations. *Biotechnology Advances* 17, 29-48

Lee SY (1996) High cell-density culture of *Escherichia coli*. *Trends in Biotechnology* 14, 98-105

Leiting B, Marsilio F, and Oçonnell JF (1998) Predictable deuteration of recombinant proteins expressed in *Escherichia coli*. *Analytical Biochemistry* 265, 351-355

Li C, Schwabe JW, Banayo E, Evans RM (1997) Coexpression of nuclear receptor partners increases their solubility and biological activities. *Proceedings of The National Academy of Sciences USA* 94, 2278-2283

Lian L-Y and Middleton DA (2001) Labelling approaches for protein structural studies by solution-state and solid-state NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy* 39, 171-190

Link AJ, Phillips D, and Church GM (1997) Methods for generating precise deletions and insertions in the genome of wild-type *Escherichia coli*: application to open reading frame characterization. *Journal of Bacteriology* 179, 6228-6237

Love CA, Lilley PE, Dixon NE (1996) Stable high-copy-number bacteriophage lambda promoter vectors for overproduction of proteins in *Escherichia coli*. *Gene* 176, 49-53

Luo Q, Shen YL, Wei DZ, Cao W (2006) Optimization of culture on the overproduction of TRAIL in high-cell-density culture by recombinant *Escherichia coli*. *Applied Microbiology and Biotechnology* 71, 184-191

Makrides S (1996) Strategies for Achieving High-Level Expression of Genes in *Escherichia coli*. *Microbiological Reviews* 60, 512-538

Marley J, Lu M, Bracken C (2001) A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolecular NMR* 20, 71-75

Menart V, Jevsevar S, Vilar M, Trobis A, Pavko A (2003) Constitutive versus thermoinducible expression of heterologous proteins in *Escherichia coli* based on strong PR, PL promoters from phage lambda. *Biotechnology and Bioengineering* 83, 181-190

Mergulhão FJM, Summers DK, Monteiro GA (2005) Recombinant protein secretion in *Escherichia coli*. *Biotechnology Advances* 23, 177-202

Miroux B, Walker JE (1996) Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology* 260, 289-298

Møller-Jensen J, Jensen RB and Gerdes K (2000) Plasmid and chromosome segregation in procaryotes. *Trends in Microbiology* 8, 313-320

Monod J (1950) La technique de culture continue: theorie et applications. *Annales de l'Institute Pasteur* 79, 390-410

Moore PB (1979) The preparation of deuterated ribosomal materials for neutron scattering. *Methods in Enzymology* 59, 639-649

Morgan WD, Kragt A and Feeney J (2000) Expression of deuterium-isotope-labelled protein in the yeast *Pichia pastoris* for NMR studies. *Journal of Biomolecular NMR* 17, 337-347

Muchmore DC, McIntosh LP, Russell CB, Anderson DE, Dahlquist FW (1989) Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. *Methods in Enzymology* 177, 44-73

Müller D, Bayer K and Mattanovich D (2006) Potentials and limitations of prokaryotic and eukaryotic expression systems for recombinant protein production – a comparative view. *Microbial Cell Factories* 5(Suppl 1): 61

Nakano K, Rischke M, Sato S, Markl H (1997) Influence of acetic acid on the growth of *Escherichia coli* K12 during high-cell-density cultivation in a dialysis reactor. *Applied Microbiology and Biotechnology* 48, 597-601

Nakano H, Shinbata T, Okumura R, Sekiguchi S, Fujishiro M, Yamane T (1999) Efficient coupled transcription/translation from PCR template by a hollow-fiber membrane bioreactor. *Biotechnology and Bioengineering* 64, 194-199

Neidhardt FC, Ingraham JL and Schaechter M (1990) Physiology of the Bacterial Cell. A Molecular Approach. Sinauer Associates, Inc., Sunderland, MA.

Neu HC, Heppel LA (1965) The release of enzymes from *Escherichia coli* by osmotic shock and during the formation of spheroplasts. *The Journal of Biological Chemistry* 240, 3685-3692

Newbury SF, Smith NH, Robinson EC, Hiles ID, Higgins CF (1987) Stabilization of translationally active mRNA by prokaryotic REP sequences. *Cell* 48, 297-310

Nietlispach D (2007) The use of deuteration for the structural study of larger proteins EMBO Practical Course: Structure determination of biological macromolecules by solution NMR July 6-13th

Nishihara K, Kanemori M, Yanagi H, and Yura T (2000) Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*. *Applied and Environmental Microbiology* 66, 884-889

Novick A and Szilard L (1950) Experiments with the chemostat on spontaneous mutations of bacteria. *Proceedings of The National Academy of Sciences USA* 36, 708-719

Paalme T, Tiisma K, Kahru A, Vanatalu K, and Vilu R (1990) Glucose limited fed-batch cultivation of *Escherichia coli* with computer-controlled fixed growth rate. *Biotechnology and Bioengineering* 35, 312-319

Paalme T and Vilu R (1992) A new method of continuous cultivation with computer-controlled change of dilution rate. In: Karim MN and Stephanopoulos G (eds) Modeling and control of biotechnical processes. Pergamon Press, 299-302

Paalme T, Kahru A, Elken R, Vanatalu K, Tiisma K and Vilu R (1995) The computer-controlled continuous culture of *Escherichia coli* with smooth change of dilution rate (A-stat). *Journal of Microbiological Methods* 24, 145-153

Paalme T, Elken R, Kahru A, Vanatalu K, Vilu R (1997) The growth-rate control in *Escherichia coli* at near to maximum growth rates: the A-stat approach. *Antonie van Leeuwenhoek* 71, 217-230

Paliy O, Bloor D, Brockwell D, Gilbert P, Barber J (2003) Improved methods of cultivation and production of deuteriated proteins from *E. coli* strains grown on fully deuteriated minimal medium. *Journal of Applied Microbiology* 94, 580-586

Patzelt H, Goto N, Iwai H, Lundstrom K, and Fernholz E (2002) Modern methods for the expression of proteins in isotopically enriched form. In: Zerbe O (ed) BioNMR in Drug Research. Wiley-VCH Verlag GmbH & Co, 1-38

Paulsson J and Ehrenberg M (1998) Trade-off between segregational stability and metabolic burden: a mathematical model of plasmid ColE1 replication control. *Journal of Molecular Biology* 279, 73-88

Pecota DC, Kim CS, Wu K, Gerdes K and Wood TK (1997) Combining the hok/sok, parDE, and pnd postsegregational killer loci to enhance plasmid stability. *Applied and Environmental Microbiology*, 1917-1924

Pervushin K, Riek R, Wider G and Wüthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of The National Academy of Sciences USA* 94, 12366-12371

Phue J-N, Shiloach J (2004) Transcription levels of key metabolic genes are the cause for different glucose utilization pathways in *E. coli* B (BL21) and *E. coli* K (JM109). *Journal of Biotechnology* 109, 21-30

Phue J-N, Noronha SB, Bhattacharyya R, Wolfe AJ, Shiloach J (2005) Glucose metabolism at high density growth of *E. coli* B and *E. coli* K: differences in metabolic pathways are responsible for efficient glucose utilization in *E. coli* B as determined by microarrays and Northern blot analyses. *Biotechnology and Bioengineering* 90, 805-820

Pirt SJ (1965) The maintenance energy of bacteria in growing cultures. *Proceedings of the Royal Society of London Series B, Biological Sciences* 163, 224-231

Poitras, P (1994) Motilin. In: Walsh JH and Dockray GJ (eds) Gut Peptides: Biochemistry and Physiology. Raven Press, New York, 261-304

Poole ES, Brown CM, and Tate WP (1995) The identity of the base following the stop codon determines the efficiency of in vivo translational termination in *Escherichia coli The EMBO Journal* 14, 151-158

Qiu J, Swartz JR, Georgiou G (1998) Expression of active human tissue-type plasminogen activator in *Escherichia coli*. *Applied and Environmental Microbiology* 64, 4891-4896

Reilly D, Fairbrother WJ (1994) A novel isotope labeling protocol for bacterially expressed proteins. *Journal of Biomolecular NMR* 4, 459-462

Riesenberg D and Guthke R (1999) High-cell-density cultivation of microorganisms. *Applied Microbiology and Biotechnology* 51, 422-430

Riesenberg D, Schulz V, Knorre WA, Pohl HD, Korz D, Sanders EA, Ross A and Deckwer W-D (1991) High cell density cultivation of *Escherichia coli* at controlled specific growth rate. *Journal of Biotechnology* 20, 17-27

Rössle M, Manakova E, Holzinger J, Vanatalu K, May RP , Heumann H (2000) Time-resolved small-angle neutron scattering of proteins in solution. *Physica B* 276-278, 532-533

Rowe DCD and Summers DK (1999) The quiescent-cell expression system for protein synthesis in *Escherichia coli Applied and Environmental Microbiology* 65, 2710-2715

Rule GS and Hitchens TK (eds) (2006) Fundamentals of protein NMR spectroscopy. Springer

Ryan W, Collier P, Loredo L, Pope J, Sachdev R (1996) Growth kinetics of *Escherichia coli* and expression of a recombinant protein and its isoforms under heat shock conditions. *Biotechnology Progress* 12, 596-601

Sambrook J, Fritsch EF and Maniatis T (1989) Molecular cloning: A laboratory manual, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 17.32-17.33

Sandén AM, Prytz I, Tubulekas I, Förberg C, Le H, Hektor A, Neubauer P, Pragai Z, Harwood C, Ward A, Picon A, Teixeira de Mattos J, Postma P, Farewell A, Nyström T, Reeh S, Pedersen S, Larsson G (2003) Limiting factors in *Escherichia coli* fed-batch production of recombinant proteins. *Biotechnology and Bioengineering* 81, 158-166

Sandman K, Grayling RA and Reeve JN (1995) Improved N-terminal processing of recombinant proteins synthesized in *Escherichia coli*. *Bio/Technology* 13, 504-506

Senn H, Lendenmann U, Snozzi M, Hamer G, Egli T (1994) The growth of *Escherichia coli* in glucose-limited chemostat cultures: a re-examination of the kinetics. *Biochimica et Biophysica Acta* 1201, 424-436

Shiloach J and Fass R (2005) Growing *E. coli* to high cell density—A historical perspective on method development. *Biotechnology Advances* 23, 345-357

Shokri A, Sandén AM, Larsson G (2003) Cell and process design for targeting of recombinant protein into the culture medium of *Escherichia coli*. *Applied Microbiology and Biotechnology* 60, 654-664

Siemeister G, Schnurr B, Mohrs K, Schachtele C, Marme D, Martiny-Baron G (1996) Expression of biologically active isoforms of the tumor angiogenesis factor VEGF in *E. coli*. *Biochemical and Biophysical Research Communications* 222, 249-255

Singh SM and Panda AK (2005) Solubilization and refolding of bacterial inclusion body proteins. *Journal of Bioscience and Bioengineering* 99, 303-310

Small P, Blankenhorn D, Welty D, Zinser E, Slonczewski JL (1994) Acid and base resistance in *Escherichia coli* and *Shigella flexneri*: role of rpoS and growth pH. *Journal of Bacteriology* 176,1729-1737

Smith HO, Wilcox KWA (1970) Restriction enzyme from *Haemophilus influenzae*. I. Purification and general properties. *Journal of Molecular Biology* 51, 379-391

Sørensen HP, Mortensen KK (2005) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *Journal of Biotechnology* 115, 113-128

Stewart EJ, Åslund F and Beckwith J (1998) Disulfide bond formation in the *Escherichia coli* cytoplasm: an in vivo role reversal for the thioredoxins. *The EMBO Journal* 17, 5543-5550

Storrs SB, Przybycien TM (1991) Commercial-scale refolding of recombinant methionyl bovine somatotropin. In: Georgiou G, DeBernardez-Clark E (eds) Protein Refolding. Washington, DC: American Chemical Society, 197-205

Studier FW (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *Journal of Molecular Biology* 219, 37-44

Studier FW, Moffatt BA (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology* 189, 113-130

Summers DK, Beton CW, Withers HL (1993) Multicopy plasmid instability: the dimer catastrophe hypothesis. *Molecular Microbiology* 8, 1031-1038

Summers D (1998) Timing, self-control and a sense of direction are the secrets of multicopy plasmid stability. *Molecular Microbiology* 29, 1137-1145

Summers DK (1991) The kinetics of plasmid loss. *Trends in Biotechnology* 9, 273-278

Summers, DK and Sherratt DJ (1984) Multimerization of high copy number plasmids causes instability: ColE1 encodes a determinant essential for plasmid monomerization and stability. *Cell* 36, 1097-1103

Swartz JR (2001) Advances in *Escherichia coli* production of therapeutic proteins. *Current Opinion in Biotechnology* 12, 195-201

Terpe K (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology* 60, 523-533

Terpe K (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology* 72, 211-222

Tomizawa J (1990) Control of ColE1 plasmid replication. Interaction of Rom protein with an unstable complex formed by RNA I and RNA II. *Journal of Molecular Biology* 212, 695-708

Torizawa T, Shimizu M, Taoka M, Miyano H, Kainosho M (2004) Efficient production of isotopically labeled proteins by cell-free synthesis: a practical protocol. *Journal of Biomolecular NMR* 30, 311-325

Tomson K, Vanatalu K (1997) Cultivation of recombinant *Escherichia coli* on deuterated medium. Abstracts of 8th European Congress on Biotechnology, Aug 17-21, Budapest, Hungary, 403

Tong L, Lin Q, Wong WKR, Ali A, Lim D, Sung WL, Hew CL, Yang DSC (2000) Extracellular expression, purification, and characterization of a winter flounder antifreeze polypeptide form *Escherichia coli*. *Protein Expression and Purification* 18, 175-181

Van de Walle M and Shiloach J (1998) Proposed mechanism of acetate accumulation in two recombinant *Escherichia coli* strains during high-density fermentation. *Biotechnology and Bioengineering* 57, 71-78

Van der Wal FJ, Koningstein G, ten Hagen CM, Oudega B, Luirink J (1998) Optimization of bacteriocin release protein (BRP)-mediated protein release by *Escherichia coli*: random mutagenesis of the pCloDF13-derived BRP gene to uncouple lethality and quasi-lysis from protein release. *Applied and Environmental Microbiology* 64, 392-398

Vanatalu K, Paalme T, Vilu R, Burkhardt N, Jünemann R, May R, Rühl M, Wadzack J, Nierhaus KH (1993) Large-scale preparation of fully deuterated cell components. Ribosomes from *Escherichia coli* with high biological activity. *European Journal of Biochemistry* 216, 315-321

Venters RA, Calderone TL, Spicer LD, Fierke CA (1991) Uniform ^{13}C isotope labeling of proteins with sodium acetate for NMR studies: application to human carbonic anhydrase II. *Biochemistry* 30, 4491-4494

Venters RA, Huang CC, Farmer BT 2nd, Trolard R, Spicer LD, Fierke CA (1995) High-level $^{2}H/^{13}C/^{15}N$ labeling of proteins for NMR studies. *Journal of Biomolecular NMR* 5, 339-344

Vethanayagam JGG and Flower AM (2005) Decreased gene expression from T7 promoters may be due to impaired production of active T7 RNA polymerase. *Microbial Cell Factories* 4: 3

Vinter T, Paalme T and Vilu R (1992) "FermExpert" - an expert system for studies and optimization of processes of microbial synthesis. In: Karim MN and Stephanopolous G

(eds) Modelling and Control in Biotechnical Processes. Pergamon Press Ltd., Headington Hill Hall, Oxford, England 467-470

Wacker M, Linton D, Hitchen PG, Nita-Lazar M, Haslam SM, North SJ, Panico M, Morris HR, Dell A, Wren BW, Aebi M (2002) N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* 298, 1790-1793

Walsh G (2006) Biopharmaceutical benchmarks 2006. *Nature Biotechnology* 24, 769-776

Wick LM, Egli T (2004) Molecular components of physiological stress responses in *Escherichia coli*. *Advances in Biochemical Engineering/Biotechnology* 89, 1-45

Wider G (2000) Structure determination of biological macromolecules in solution using nuclear magnetic resonance spectroscopy. *Biotechniques* 29, 1278-1294

Wild J, Szybalski W (2004) Copy-control tightly regulated expression vectors based on pBAC/oriV. *Methods in Molecular Biology* 267, 155-167

Wilkinson DL, Harrison RG (1991) Predicting the solubility of recombinant proteins in *Escherichia coli*. *Biotechnology (NY)* 9, 443-448

Williams SG, Cranenburgh RM, Weiss AME, Wrighton CJ, Sherratt DJ and Hanak JAJ (1998) Repressor titration: a novel system for selection and stable maintenance of recombinant plasmids. *Nucleic Acids Research* 26, 2120-2124

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

Wong HH, Kim YC, Lee SY, Chang HN (1998) Effect of post-induction nutrient feeding strategies on the production of bioadhesive protein in *Escherichia coli*. *Biotechnology and Bioengineering* 60, 271-276

Yavachev L, Ivanov I (1988) What does the homology between *E. coli* tRNAs and RNAs controlling ColE1 plasmid replication mean? *Journal of Theoretical Biology* 131, 235-241

Yee L and Blanch HW (1992) Recombinant protein expression in high cell density fed-batch cultures of *Escherichia coli*. *Bio/Technology* 10, 1550-1556

Yu, H (1999) Extending the size limit of protein nuclear magnetic resonance. *Proceedings of The National Academy of Sciences USA* 96, 332-334

Zahn K (1996) Overexpression of an mRNA dependent on rare codons inhibits protein synthesis and cell growth. Journal of Bacteriology 178, 2926-2933

Zhang Z, Gildersleeve J, Yang YY, Xu R, Loo JA, Uryu S, Wong CH, Schultz PG (2004) A new strategy for the synthesis of glycoproteins. *Science* 303, 371-373

Zhang Y, Buchholz F, Muyrers JP, Stewart AF (1998) A new logic for DNA engineering using recombination in *Escherichia coli*. *Nature Genetics* 20, 123-128

PUBLICATIONS

ARTICLE I

Tomson K, Liik P and Vilu R. Using of genetically unstable recombinant producers.

Reprinted from: *Proceedings of the Estonian Academy of Sciences, Biology,* 1989, 38, 1-8
With permission from Estonian Academy Publishers

ARTICLE II

Tomson K and Vilu R. Stability of recombinant *Escherichia coli* in continuous culture.

Reprinted from: *Proceedings of the Estonian Academy of Sciences, Biology*, 1990, 39, 93-95
With permission from Estonian Academy Publishers

ARTICLE III

Tomson K, Paalme T, Laakso PS and Vilu R. Automatic laboratory-scale fed-batch procedure for production of recombinant proteins using inducible expression systems of *Escherichia coli*.

Reprinted from: *Biotechnology techniques*, 1995, 9, 793-798
With permission from Springer

ARTICLE IV

Vanatalu K, Tomson K, Paalme T, Tiisma K and Vilu R. Application of dense culture for *in vivo* deuteration of microorganisms.

Reprinted from: *Proceedings* of the Workshop on Deuteration of Biological Molecules for Structural and Dynamic Studies. Application to Neutron Scattering and NMR. May 19-25, 1998, Dubna, Russia, 7-12

ARTICLE V

Tomson K, Barber J, Vanatalu K. Adaptastat - a new method for optimising of bacterial growth conditions in continuous culture: Interactive substrate limitation based on dissolved oxygen measurement.

Reprinted from: *Journal of Microbiological Methods*, 2006, 64, 380-390, Epub 2005 Jul 1
With permission from Elsevier

ARTICLE VI

Patent application: Method of cultivation and recombinant protein expression in *Escherichia coli*; Owner: Tallinn University of Technology; Authors: Kalju Vanatalu, Katrin Tomson; Priority number: US60/ 827, 336; Priority date: 29.09.2006

ARTICLE VII

Massad T, Jarvet J, Tanner R, Tomson K, Smirnova J, Palumaa P, Sugai M, Kohno T, Vanatalu K, Damberg P. Maximum entropy reconstruction of joint phi, psi-distribution with a coil-library prior: the backbone conformation of the peptide hormone motilin in aqueous solution from phi and psi-dependent J-couplings.

Reprinted from: *Journal of Bimolecular NMR*, 2007, 38, 107-123, Epub 2007 Apr 26

ABSTRACT

Production of labelled recombinant proteins in fed-batch systems in *Escherichia coli*

Despite remarkable progress in gene expression studies and an increasing number of commercially available expression systems, development of a production process for a particular recombinant protein remains complicated and requires trial, experience and optimisation.

This study investigates a variety of parameters (e.g. culture stability, selective pressure, temperature, metabolic by-products, cultivation methods, composition and deuteration of growth media) that influence the production and labelling of recombinant proteins.

The stability of expression was studied and induction conditions were optimised for several unlabelled recombinant proteins (rat and human catechol-*O*-methyltransferase, calf prochymosin, human troponin C, yeast ubiquitin hydrolase) and $^{15}N/\ ^{2}H/\ ^{13}C$- labelled proteins (sperm whale myoglobin, *E. coli* GroEL and GroES, *E. coli* ribosomal protein L29, GST-GFPuv, yeast ubiquitin-human motilin) in fed-batch cultivation systems. A feed-forward procedure that enables the cultivation of cells at desired growth rate and a novel adaptive feeding algorithm (named adaptastat) were developed to produce labelled recombinant proteins and biomasses in high cell density culture.

The expression and purification of ^{15}N-, and (double-labelled) $^{2}H^{15}N$-, $^{13}C^{15}N$-motilin (a peptide hormone), as a fusion protein was investigated as a model case for the production of various labelled polypeptide and protein molecules.

The purified recombinant proteins from this work were highly enriched and pure. Importantly, they were successfully used by our collaboration partners for structural and functional studies using NMR, neutron scattering, and other techniques.

Keywords: recombinant proteins, peptides, *Escherichia coli*, stability, selective pressure, stress, production, fed-batch, automatic, adaptastat, continuous cultivation, isotopic labelling, specific growth rate, NMR, neutron scattering.

KOKKUVÕTE

Märgistatud rekombinantsete valkude tootmine *Escherichia coli* fed-batch süsteemides

Vaatamata intensiivsele geeniekspressiooni alasele uurimistööle ja paljude (sh kommertsiaalsete) ekspressioonisüsteemide kättesaadavusele on iga konkreetse valgu tootmisprotseduuri väljatöötamine keeruline ülesanne, mis baseerub paljuski kogemustele ja katse-eksituse meetodile.

Kirjanduse ülevaates käsitletakse olulisi etappe rekombinantse valgu tootmisel: ekspressioonisüsteemi valikut ja kloneerimist, rekombinantsete tüvede kultiveerimist ja valgu ekspressiooni ning valgu puhastamist. Samuti antakse ülevaade rekombinantse valgu saagist ja kvaliteeti mõjutavatest teguritest, nende tootmise hetkeseisust, *E. coli* ekspressioonisüsteemide potentsiaalist ja kohast teiste süsteemide hulgas ning valkude märgistamisest TMR uuringuteks.

Eksperimentaalse töö eesmärgiks oli uurida ekspressioonisüsteemide ja kultiveerimistingimuste mõju rekombinantsete valkude ekspressioonile, optimeerida vastavad tootmisprotsessid ja toota rekombinantseid biomasse/ valke uurimistöödeks vajalikes kogustes. Selleks uuriti mitmete kasvu- ja ekspressiooni-tingimuste nagu temperatuuri, söötme koostise ja deutereerimise, erinevate süsinikuallikate, ainevahetuse kõrvalproduktide tekke, indutseerimistiheduse, induktori (IPTG) kontsentratsiooni, kultiveerimismeetodite, jne. mõju rekombinantsete valkude ekspressiooni stabiilsusele *E. coli* kultuuris.

^{15}N/ ^{13}C/ ^{2}H-märgistatud valkude tootmiseks uuriti lisaks eelpool loetletutele veel *E. coli* kultuuride kasvu ja ekspressiooni hinnalt soodsamatel ent raskemini omastatavatel süsinikuallikatel (suktsinaadil, atsetaadil) ja deutereeritud kasvukeskkonnas.

Tiheda *E. coli* kultuuri saamiseks töötati välja nii ettemääratud söötmisprofiiliga (feed-forward) ja uudne adaptiivne (adaptastaat) fed-batch kultiveerimismeetod. Näidati, et tihedas kultuuris õnnestub saavutada rekombinantsete valkude (roti ja inimese katehhool-O-metüültransferaas, vasika prokümosiin, inimese troponiin C; ^{13}C/ ^{2}H/ ^{15}N-märgistatud valgud: kaseloti müoglobiin, *E. coli* chaperon-valgud GroEL ja GroES, *E. coli* ribosoomivalk L29, reportervalk GST-GFPuv, inimese peptiidhormoon motiliin liitvalgu koosseisus) standardmeetoditega võrreldav ekspressioonitase ja kõrge mahuline tootlikkus. Deutereeritud ühendite korral õnnestus seeläbi vähendada tootmiskulusid kümneid kordi.

Töös käsitletakse näitena stabiilsete isotoopidega märgistatud motiliini mitmeetapilise tootmisprotseduuri väljatöötamist: biomassi kasvatamise, rakkude lüüsimise, afiinsuskromatograafia, motiliini ensümaatilise liitvalgu küljest lahtilõikamise ja motiliini täiendava puhastamise (HPLC) optimeerimist.

Töö tulemusena toodeti kõrge isotoopse rikastusega ja puhtusastmega ^{15}N-, ^{2}H^{15}N- ja ^{13}C^{15}N-märgistatud motiliinid ning rida märgistamata ja märgistatud

biomasse teiste rekombinantsete valkude puhastamiseks, mida koostööpartnerid kaasutasid erinevates struktuuri ja funktsionaalsuse uuringutes.

Märksõnad: rekombinantsed valgud, peptiidid, *Escherichia coli*, stabiilsus, selektiivne press, stress, tootmine, *fed-batch*, automaatne, adaptastaat, pidevkultiveerimine, isotoop-märgistamine, kasvu erikiirus, NMR, neutronhajumine.

DISSERTATIONS DEFENDED AT TALLINN UNIVERSITY OF TECHNOLOGY ON *NATURAL AND EXACT SCIENCES*

1. **Olav Kongas**. Nonlinear dynamics in modeling cardiac arrhytmias. 1998.

2. **Kalju Vanatalu**. Optimization of processes of microbial biosynthesis of isotopically labeled biomolecules and their complexes. 1999.

3. **Ahto Buldas**. An algebraic approach to the structure of graphs. 1999.

4. **Monika Drews**. A metabolic study of insect cells in batch and continuous culture: application of chemostat and turbidostat to the production of recombinant proteins. 1999.

5. **Eola Valdre**. Endothelial-specific regulation of vessel formation: role of receptor tyrosine kinases. 2000.

6. **Kalju Lott**. Doping and defect thermodynamic equilibrium in ZnS. 2000.

7. **Reet Koljak**. Novel fatty acid dioxygenases from the corals *Plexaura homomalla* and *Gersemia fruticosa.* 2001.

8. **Anne Paju**. Asymmetric oxidation of prochiral and racemic ketones by using sharpless catalyst. 2001.

9. **Marko Vendelin**. Cardiac mechanoenergetics *in silico*. 2001.

10. **Pearu Peterson**. Multi-soliton interactions and the inverse problem of wave crest. 2001.

11. **Anne Menert**. Microcalorimetry of anaerobic digestion. 2001.

12. **Toomas Tiivel**. The role of the mitochondrial outer membrane in *in vivo* regulation of respiration in normal heart and skeletal muscle cell. 2002.

13. **Olle Hints**. Ordovician scolecodonts of Estonia and neighbouring areas: taxonomy, distribution, palaeoecology, and application. 2002.

14. **Jaak Nõlvak**. Chitinozoan biostratigrapy in the Ordovician of Baltoscandia. 2002.

15. **Liivi Kluge**. On algebraic structure of pre-operad. 2002.

16. **Jaanus Lass**. Biosignal interpretation: Study of cardiac arrhytmias and electromagnetic field effects on human nervous system. 2002.

17. **Janek Peterson**. Synthesis, structural characterization and modification of PAMAM dendrimers. 2002.

18. **Merike Vaher**. Room temperature ionic liquids as background electrolyte additives in capillary electrophoresis. 2002.

19. **Valdek Mikli**. Electron microscopy and image analysis study of powdered hardmetal materials and optoelectronic thin films. 2003.

20. **Mart Viljus**. The microstructure and properties of fine-grained cermets. 2003.

21. **Signe Kask**. Identification and characterization of dairy-related *Lactobacillus.* 2003.

22. **Tiiu-Mai Laht**. Influence of microstructure of the curd on enzymatic and microbiological processes in Swiss-type cheese. 2003.

23. **Anne Kuusksalu**. 2–5A synthetase in the marine sponge *Geodia cydonium*. 2003.

24. **Sergei Bereznev**. Solar cells based on polycristalline copper-indium chalcogenides and conductive polymers. 2003.

25. **Kadri Kriis**. Asymmetric synthesis of C_2-symmetric bimorpholines and their application as chiral ligands in the transfer hydrogenation of aromatic ketones. 2004.

26. **Jekaterina Reut**. Polypyrrole coatings on conducting and insulating substracts. 2004.

27. **Sven Nõmm**. Realization and identification of discrete-time onlinear systems. 2004.

28. **Olga Kijatkina**. Deposition of copper indium disulphide films by chemical spray pyrolysis. 2004.

29. **Gert Tamberg**. On sampling operators defined by Rogosinski, Hann and Blackman windows. 2004.

30. **Monika Übner**. Interaction of humic substances with metal cations. 2004.

31. **Kaarel Adamberg**. Growth characteristics of non-starter lactic acid bacteria from cheese. 2004.

32. **Imre Vallikivi**. Lipase-catalysed reactions of prostaglandins. 2004.

33. **Merike Peld**. Substituted apatites as sorbents for heavy metals. 2005.

34. **Vitali Syritski**. Study of synthesis and redox switching of polypyrrole and poly(3,4-ethylenedioxythiophene) by using *in-situ* techniques. 2004.

35. **Lee Põllumaa**. Evaluation of ecotoxicological effects related to oil shale industry. 2004.

36. **Riina Aav**. Synthesis of 9,11-secosterols intermediates. 2005.

37. **Andres Braunbrück**. Wave interaction in weakly inhomogeneous materials. 2005.

38. **Robert Kitt**. Generalised scale-invariance in financial time series. 2005.

39. **Juss Pavelson**. Mesoscale physical processes and the related impact on the summer nutrient fields and phytoplankton blooms in the western Gulf of Finland. 2005.

40. **Olari Ilison**. Solitons and solitary waves in media with higher order dispersive and nonlinear effects. 2005.

41. **Maksim Säkki**. Intermittency and long-range structurization of heart rate. 2005.

42. **Enli Kiipli**. Modelling seawater chemistry of the East Baltic Basin in the late Ordovician–Early Silurian. 2005.

43. **Igor Golovtsov**. Modification of conductive properties and processability of polyparaphenylene, polypyrrole and polyaniline. 2005.

44. **Katrin Laos**. Interaction between furcellaran and the globular proteins (bovine serum albumin β–lactoglobulin). 2005.

45. **Arvo Mere**. Structural and electrical properties of spray deposited copper indium disulphide films for solar cells. 2006.

46. **Sille Ehala**. Development and application of various on- and off-line analytical methods for the analysis of bioactive compounds. 2006.

47. **Maria Kulp**. Capillary electrophoretic monitoring of biochemical reaction kinetics. 2006.

48. **Anu Aaspõllu.** Proteinases from *Vipera lebetina* snake venom affecting hemostasis. 2006.

49. **Lyudmila Chekulayeva**. Photosensitized inactivation of tumor cells by porphyrins and chlorins. 2006.

50. **Merle Uudsemaa**. Quantum-chemical modeling of solvated first row transition metal ions. 2006.

51. **Tagli Pitsi**. Nutrition situation of pre-school children in Estonia from 1995 to 2004. 2006.

52. **Angela Ivask**. Luminescent recombinant sensor bacteria for the analysis of bioavailable heavy metals. 2006.

53. **Tiina Lõugas**. Study on physico-chemical properties and some bioactive compounds of sea buckthorn (*Hippophae rhamnoides* L.). 2006.

54. **Kaja Kasemets**. Effect of changing environmental conditions on the fermentative growth of Saccharomyces cerevisae S288C: auxo-accelerostat study. 2006.

55. **Ildar Nisamedtinov**. Application of ^{13}C and fluorescence labeling in metabolic studies of Saccharomyces spp. 2006.

56. **Alar Leibak**. On additive generalisation of Voronoï's theory of perfect forms over algebraic number fields. 2006.

57. **Andri Jagomägi**. Photoluminescence of chalcopyrite tellurides. 2006.

58. **Tõnu Martma**. Application of carbon isotopes to the study of the Ordovician and Silurian of the Baltic. 2006.

59. **Marit Kauk**. Chemical composition of $CuInSe_2$ monograin powders for solar cell application. 2006.

60. **Julia Kois**. Electrochemical deposition of $CuInSe_2$ thin films for photovoltaic applications. 2006.

61. **Ilona Oja Açik**. Sol-gel deposition of titanium dioxide films. 2007.

62. **Tiia Anmann**. Integrated and organized cellular bioenergetic systems in heart and brain. 2007.

63. **Katrin Trummal**. Purification, characterization and specificity studies of metalloproteinases from *Vipera lebetina* snake venom. 2007.

64. **Gennadi Lessin**. Biochemical definition of coastal zone using numerical modeling and measurement data. 2007.

65. **Enno Pais**. Inverse problems to determine non-homogeneous degenerate memory kernels in heat flow. 2007.

66. **Maria Borissova**. Capillary electrophoresis on alkylimidazolium salts. 2007.

67. **Karin Valmsen**. Prostaglandin synthesis in the coral *Plexaura homomalla*: control of prostaglandin stereochemistry at carbon 15 by cyclooxygenases. 2007.

68. **Kristjan Piirimäe**. Long-term changes of nutrient fluxes in the drainage basin of the gulf of Finland – application of the PolFlow model. 2007.

69. **Tatjana Dedova**. Chemical spray pyrolysis deposition of zinc sulfide thin films and zinc oxide nanostructured layers. 2007.