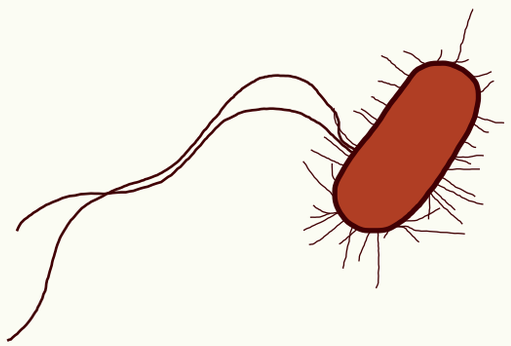
Determination of protein synthesis on a proteomic scale

Kramer, G.

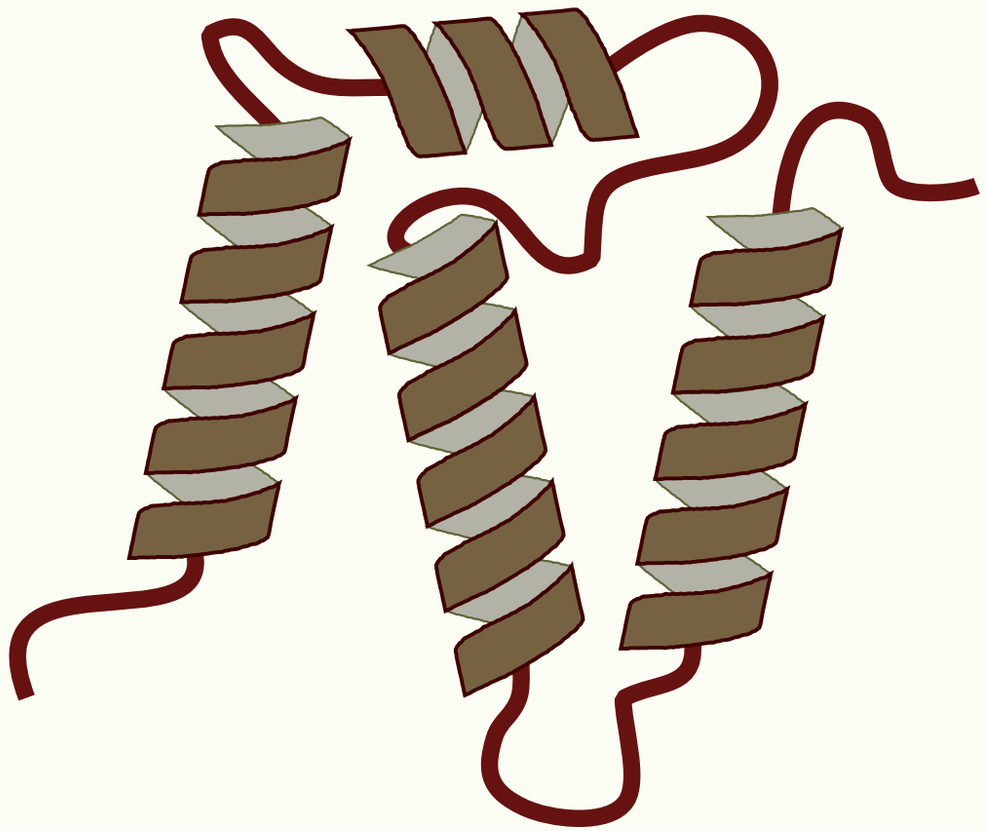
Publication date
2011
Document Version
Final published version



Determination of Protein Synthesis on a Proteomic Scale

Gertjan Kramer

Determination of Protein Synthesis on a Proteomic Scale



Gertjan Kramer

Determination of Protein Synthesis on a Proteomic Scale

Determination of Protein Synthesis on a Proteomic Scale

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties
ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op woensdag 26 oktober 2011, te 14:00 uur

door

Gertjan Kramer

geboren te Lisse

Promotor:	Prof. Dr. C.G. de Koster
Copromotores:	Dr. J.W. Back Dr. L. de Jong
Overige leden:	Prof. Dr. J.M.F.G. Aerts Prof. Dr. J. Vandekerckhove Prof. Dr. S. Brul Prof. Dr. K.J. Hellingwerf Prof. Dr. R. van Driel Prof. Dr. Ir. P. Schoenmakers Dr. J. Krijgsveld

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

voor Marcela,
amor non celatur

“Our ideas held no water,
but we used them like a dam.”

Isaac Brock

Contents

1

Regulation of protein levels: balancing synthesis and degradation

Introduction

The cell is the smallest unit of life and during evolution nature has laid down a cellular blueprint that is remarkably similar in its basic layout throughout all life-forms. In order to survive and procreate autonomously, a lipid membrane encapsulates the hereditary information and separates it from the environment. This hereditary information -the DNA- is transcribed into RNA which is in turn translated into proteins that are involved in most cellular processes that ensure survival of cellular structures against entropy and facilitate transmission of hereditary information to subsequent generations. The central dogma of biology, where information flows from DNA through RNA to proteins, has proven to be a universal concept within the whole of cellular life as well.

'omics' and post-transcriptional regulation—Advances in DNA sequencing led to the first of the 'omics' revolutions, i.e. genomics. Genomic sequencing made available the complete genetic information of increasingly complex organisms culminating in the sequencing of the human genome (1, 2). With complete genome sequences it became clear that complexity of organisms does not automatically scale with the number of protein coding genes in its genome. Thus, other explanations need to be found for differences in complexity of organisms that have similar numbers of genes within their genome. Post-transcriptional mechanisms such as alternative splicing and post-translational modification can generate a multitude of distinct gene products from a single gene, which in combination with more complex patterns of gene expression may offer an explanation for differences in phenotypic complexity. Here the second of the 'omics' fields, i.e. transcriptomics, using, amongst others, the oligonucleotide-array (or microarray) enables genome-wide measurement of gene expression patterns, through quantitation of messenger-RNA (mRNA) transcript levels in the cell (3-6).

Microarrays allow us to follow the complex changes in gene-expression patterns induced by different growth conditions, the cell cycle, stages of embryonic development, or during diseases such as cancer. In this manner much has been learned about how the cell orchestrates transcription of genes according to its needs. Through such genome-wide expression studies, in addition to much more specific biochemical and molecular biology approaches, many types of transcriptional regulation have been unravelled. A highly complex system of regulatory proteins involved in promoting and silencing transcription of genes was discovered and continues to be explored (7, 8). However, for protein coding genes the transcript is not the final goal of gene-expression. A basic, not always specifically acknowledged, premise in genome-wide measurement of transcript levels as an indicator of how the cell regulates gene expression is that protein levels closely follow transcript levels. However, apart from transcriptional regulation, there are several additional points in the road from transcript to protein where regulation can occur (Figure 1). Together these additional regulatory mechanisms are known as post-transcriptional (7, 8), providing a second layer of regulation on top of transcriptional regulation that can explain phenotypic complexity using a defined set of genes. Thus, transcript levels need not always directly reflect protein levels,

and differing degrees of correlation have been found between (changes in) cellular protein levels and (changes in) genome-wide transcript levels (9-21).

These studies into global changes of protein levels in the cell are part of the third of the 'omics' approaches: proteomics (22-25), the study of 'all' proteins expressed in the cell. In the field of proteomics, the mass spectrometer has become the analytical technique of choice for the identification and quantitation of vast numbers of proteins, often preceded by various separation techniques such as gel electrophoresis and high performance liquid chromatography to analyse the complex mixtures of molecules encountered in the cell. Mass spectrometry (MS) plays an equally central role in the fourth of the 'omics' fields i.e. metabolomics or the study of all metabolites and small molecules of the cell. With the development of these '-ome' wide approaches of different groups of molecules, attempts to integrate these vast amounts of data into models which accurately describe and predict cellular behaviour are underway, often referred to as 'systems biology' (26, 27).

To accurately describe and model cellular behaviour it is crucial that the different parts of regulatory cascades are understood so they can be accurately represented in the mathematical equations of a cellular model. For this the already substantial knowledge regarding transcriptional regulation should be supplemented by increasing knowledge about possible post-transcriptional levels of regulation during adaptation to different environmental

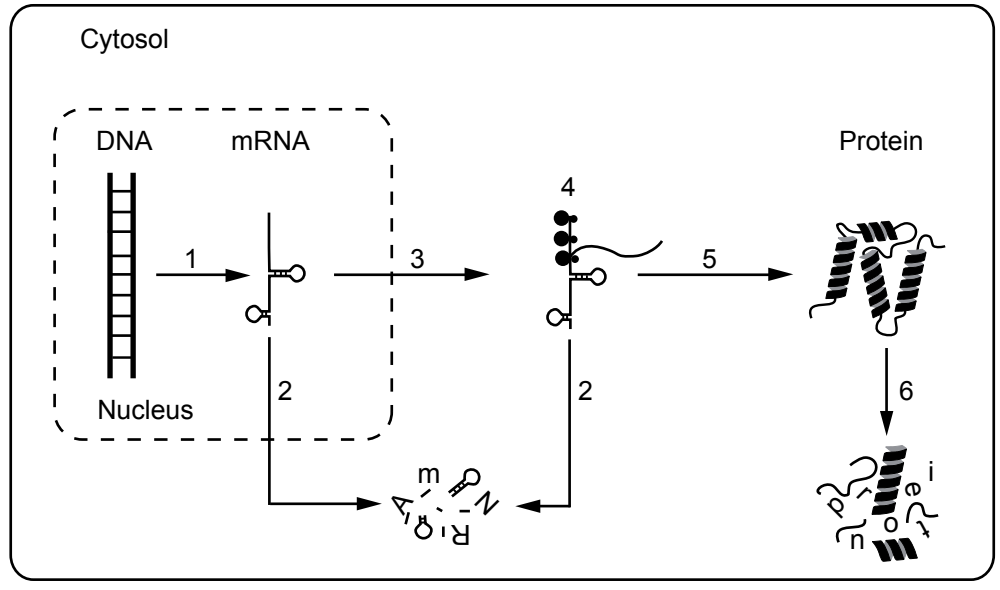


Figure 1. **From DNA to protein.** Gene expression starts with transcription (1) of genetic information, generating mRNA molecules, from which (in eukaryotes) introns are removed, a 5' cap and poly-A tail are added prior to export from the nucleus (3). mRNA levels measured in the cell are themselves a balance of the rate of transcription and degradation (2), of which the net effect is measured by microarrays between different cellular states. The exported mature mRNA is bound by ribosomes that initiate translation (4) into a polypeptide chain, which is folded and may undergo different post-translational modifications (5) before or after becoming a mature protein. Mature proteins are subject to degradation (6) and have differing half-lives.

conditions, stages of development or in health and disease. As mentioned, differing correlations have been found between transcript and protein levels in different experiments, part of which might be attributable to technical variation in the various transcriptomic and proteomic assays (15, 28, 29). Another important part is due to post-transcriptional regulation, attenuating protein level changes independently from transcription levels, many examples of which are already known from literature (see below). However, to detect and further unravel post-transcriptional regulation on a genome-wide basis is a daunting challenge at the moment. Using modern mass spectrometric techniques, it is possible to routinely quantify large numbers of protein levels inside the cell and compare these to transcript levels obtained from microarrays or, more recently, deep-sequencing techniques (30-33). But cellular levels of individual proteins and changes thereof are of course governed by two separate processes. Cellular protein levels result from a balance of protein synthesis and degradation (Figure 1). Of these, the latter is independent of the concentration of the transcript. The translation rate of a protein is not only determined by the mRNA level but also by the frequency with which ribosomes complete the synthesis of a protein per encoding mRNA molecule. This frequency can differ between different mRNAs and even for the same type of mRNA between different physiological conditions. Thus, large differences in correlation between protein and transcript levels can be caused by different rates of translation, degradation, or a combination of the two, making the identification of the mechanism(s) responsible difficult for individual proteins. As a consequence, any approach which can measure synthesis-rate as well as degradation-rate for individual proteins on a proteome-wide scale would be a very valuable addition to the proteomics toolbox.

Mechanisms and regulation of protein synthesis and degradation.

Quite a number of different processes are covered by the term post-transcriptional regulation, amongst which are mRNA-stability and -splicing as well as post-translational modifications such as glycosylation, phosphorylation, methylation etc. that affect protein conformation and function. Other regulatory events that impact protein function or enzymatic conversion rates, such as enzyme inhibition or allosteric regulation can also be considered part of post-transcriptional regulation. However, for the purpose of this thesis we are interested in post-transcriptional regulation that affects either translation or degradation (i.e. protein half-life) of proteins in order to get a better insight into how protein concentrations are regulated. As such, we leave out considerations about protein function or activity even though these are no less important for a full description of regulation of protein expression and function.

Protein synthesis (translation of the mRNA transcript into a polypeptide chain) has been studied more intensely than protein degradation and as a result much more is known about underlying regulatory mechanisms involved. This process, which is covered in basic biochemistry texts (7, 8), is the translation of the genetic information contained in the ribonucleic mRNA-molecule into the amino acid sequence of the polypeptide chain to form a protein. The information in the single-stranded mRNA is coded in triplets of ribonucleotides

which are read by the triplet anti-codon of a transfer-RNA (tRNA) by base-pairing to the mRNA (Table I). These tRNAs each carry an amino acid residue, specific for their anti-codon, ensuring the ribonucleic-acid triplets are read correctly and the correct amino acids are added to the growing polypeptide chain. Translation is performed by the ribosomes, complexes made up of RNA and proteins, which bind to the mRNA molecule upstream of an AUG start site to initiate translation. It is aided in binding by different initiation factors and subsequently the ribosome catalyzes the formation of a polypeptide chain from the amino acids linked to the tRNA, assisted by elongation factors. When the ribosome reaches a stop-codon on the mRNA molecule the polypeptide chain is released by interaction with a release factor and the ribosome is recycled for another round of translation.

TABLE I
The genetic code

1st Base	**2nd Base**								
	U			C		A		G	
U		UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
		UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
		UUA	Leu	UCA	Ser	UAA	Ochre (Stop)	UGA	Opal (Stop)
		UUG	Leu	UCG	Ser	UAG	Amber (Stop)	UGG	Trp
C		CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
		CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
		CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
		CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
A		AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
		AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
		AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
		AUG	Met (Start)	ACG	Thr	AAG	Lys	AGG	Arg
G		GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
		GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
		GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
		GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

Regulation of translation rate—from this very simplified description of translation of transcripts into proteins it is already clear that there are several points at which the rate of translation of a transcript can be influenced independently of transcript abundance (28, 34-39). When an mRNA molecule is ready for translation, both in eu- and prokaryotes, translation initiation is the first point at which regulation may take place. Initiation is aided by initiation factors and phosphorylation of these can modulate translation rate in eukaryotes (36, 40, 41). In prokaryotes the translation initiation site is preceded by the Shine Dalgarno sequence (42) which is complementary to the 3' end of 16S rRNA, and mediates 30S ribosome binding near the first AUG-codon. Shine Dalgarno sequences with lower complementarity to rRNA result

in a lower translation initiation efficiency of the transcript (43-45). In eukaryotes the Kozak sequence (46) plays a similar role. mRNA transcripts are usually not in a linear conformation, i.e. they may form hairpin-like structures which can inhibit initiation complex binding to the transcript. These hairpin secondary structures can be used to influence ribosome binding in different ways. For instance, binding of regulatory proteins or ribosomes on other parts of the transcript may change the structure to allow ribosome binding (7, 47). Alternatively, environmental conditions such as changes in temperature may affect secondary structure of a transcript as well. In *Escherichia coli* this is used as a temperature sensor, by which a change in conformation of the *rpoH* transcript at higher temperatures permits translation of the heat shock sigma factor, σ^{32} (48-50). Proteins that bind to translation initiation sites, or the upstream region, can also prevent binding of ribosomes to the transcript to initiate translation (51). Elegant examples of this are found with ribosomal proteins binding to their own mRNA-transcripts whenever free rRNA is not available. This feedback mechanism prevents translation of excess ribosomal proteins (52-55).

Not only proteins bind to mRNA transcripts. Small RNA transcripts that do not code for proteins can, through (imperfect) base-pairing, influence their translation as well. Known as small RNAs (sRNAs) in prokaryotes (56, 57) or micro-RNAs (miRNAs) in eukaryotes (58, 59), these RNAs impact translation of transcripts in various ways. On the one hand, they can act inhibitory by either preventing ribosome binding to mRNA, or by blocking translation initiation. On the other hand, sRNAs can release inhibitory secondary structures elsewhere in the mRNA molecule, thereby stimulating translation. Furthermore, sRNAs can influence transcript stability, by either shielding or promoting access to RNase cleavage sites (56-62).

Apart from initiation, the rate at which ribosomes read through a transcript (elongation) can also be influenced by mRNA secondary structure, though bound ribosomes can read through hairpin-loops and are usually only slowed, not stalled, by such structures (39, 63-65). Codon bias is another feature known to influence translation rate (66, 67), in line with the observation that different tRNA species abundances correlate with the frequency of use of their cognate codons in the genome. Furthermore, highly expressed genes have a strong bias toward use of more frequent codons (28, 68-70). Phosphorylation of eukaryotic elongation factors can influence the rate of translation as well (71, 72). The identity of the stop-codon and its surrounding sequence also have been found to influence translation termination efficiency (73-77).

The mechanisms which influence initiation, elongation and termination of translation correlate with ribosome occupancy of an mRNA species. Multiple ribosomes can occupy and translate the same transcript, which is an indication of how actively a transcript is being translated. This can be assayed by isolation of actively translating ribosome-mRNA complexes (polysomes) by density centrifugation or immuno-precipitation and quantifying attached transcripts by microarray (78-80) or deep-sequencing of ribosomal footprints (81). Levels of transcripts attached to polysomes were found to correlate better with protein abundances than transcript abundance alone (11, 81), showing that translational regulation

exerts significant influence. All in all, the above shows some different regulatory mechanisms that can account for discrepancies between translation rate and transcript abundance, with more possibly awaiting discovery.

Regulation of protein half-life—When correlations between mRNA transcripts and protein levels are investigated, the regulation of protein synthesis-rate is only part of the equation, as protein levels are regulated by degradation at least as extensively. Regulation and mechanisms of protein degradation are less well studied than those of translation, but (partial) degradation plays a role in protein 'maturation' and activity regulation, such as with co-translational removal of the N-terminal methionine or the signal sequence of membrane proteins. Degradation removes misfolded, damaged or otherwise aberrant proteins that could harm the cell and removes proteins at the end of their life-cycle. How degradation of various proteins is regulated, and how the half-life of individual proteins is determined is an important aspect of understanding how synthesis and degradation are balanced to change protein levels with respect to cellular needs.

Protein degradation is carried out by a variety of proteases, many of which form complexes. In eukaryotes the 26S proteasome is the major complex involved, and some close homologues of it are found in several prokaryotes (82-86). The 20S subunit of the 26S proteasome is a multi-subunit protease with its proteolytic active site confined to the internal compartment of the protein complex. In order for proteins to be degraded they have to enter the inner proteolytic-compartment, for which unfolding is necessary. Here the 19S subunit of the 26S proteasome comes in to play, recognizing proteins labelled for degradation, and unfolding them in an ATP-dependant manner. Inside the proteolytic compartment unfolded proteins are degraded into peptides, which exit the complex for further degradation into amino acids (87-89). *E. coli* has its own complement of self assembling protease complexes e.g. HslU/V and ClpA/P, as well as a variety of other ATP-dependent and -independent proteases (90). In eukaryotes proteolysis also takes place in the lysosome, a small membrane enclosed organelle which carries out various functions in catabolism. Whole sections of the cytosol, including organelles, are internalized in autophagic vacuoles and degraded after fusion with lysosomes under nutrient deprived conditions in a process called macro-autophagy (91). During normal conditions aspecific turnover occurs by invagination of small sections of the cytosol into the lysosome by micro-autophagy. Both processes are aspecific, but a third pathway called chaperone mediated autophagy translocates specific proteins across the lysosomal membrane under nutrient deprivation conditions and provides targeted degradation of its substrate proteins (92). Apart from these lysosomal pathways for the degradation of cytosolic proteins, extracellular and membrane proteins are degraded by endocytosis and secretory proteins by crinophagy, both involving vesicle fusion with the lysosomes, analogous to macro-autophagy.

As mentioned above, proteins need to be unfolded to enter the proteolytic core of the protease complexes and here the non-protease (chaperone) subunits play a role in recognition

and unfolding. What do these proteins recognize on their targets? The most well known feature of proteins influencing their stability is the N-terminal amino acid residue. It was established by comparing LacZ with different N-terminal residues that the half-life of the protein was strongly influenced, which is known as the N-end rule (93-96). The N-termini are recognized by ubiquitin-ligating enzymes in eukaryotes, which add poly-ubiquitin chains to lysines, which are in turn recognized by the 19S subunit of the proteasome as a mark for degradation (97, 98). In *E. coli* the ClpA/P protease complex was found to directly mediate N-end rule degradation (99). It is noteworthy that although the protease involved was known through experiments with artificial substrates, only recently endogenous proteins were identified in *E. coli* that are degraded in an N-end dependent manner (100). All proteins start with methionine (or with formyl-methionine in bacteria) during translation, which is a stabilizing residue, in principle making all proteins 'stable'. However, methionine is often removed co-translationally revealing a new N-terminus (101), while specific internal cleavage of a protein during maturation (e.g. removal of a signal sequence) can also change the N-terminal residue, strongly influencing the mature protein's half-life. Apart from the N-end rule a 'PEST' sequence motif is found in proteins with a short half-life in eukaryotes (102).

C-terminal sequences can also influence degradation or specifically target proteins for degradation. An example in *E. coli* is the rescue of stalled ribosomes that reach the 3' end of an mRNA molecule before a stop codon is reached. To release the ribosome and make it available for further rounds of translation, ssrA, a stable 362 base RNA which folds into a tRNA like structure at its 3' end binds to the A-site of the ribosome. ssrA is charged with alanine and the ribosome continues translation of the ssrA message adding a specific C-terminal tag (AANDENYLALAA) marking the incomplete polypeptide for rapid degradation (7). Signal sequences also play a role in lysosomal targeting of specific proteins under starvation conditions. 'KFERQ' was characterized as the signal sequence that targets specific proteins for chaperone mediated lysosomal degradation (92).

Some protease target sequences normally are obscured, but when exposed upon unfolding, promote degradation of unfolded proteins. Removal of aberrant or unfolded proteins is an important role for proteases and it is not coincidental that a number of proteases in *E. coli* are known as heat shock proteins owing to the conditions under which they were first discovered. Unfolded proteins are recognized by different chaperones that attempt to refold them but can also deliver them to proteases for subsequent degradation (48, 103-105). Regulation of protein half-life can be an important mechanism. During heat shock in *E. coli* for instance, apart from translational control of σ^{32}, the level of this σ-factor is also controlled via its half-life. During growth at normal temperatures, σ^{32} has a half-life of only a few minutes, effected by some of the proteases mentioned above. However, at higher temperatures stability of σ^{32} rapidly increases (106-108), thus raising the levels in concert with the translational activation. Elevated levels of this heat shock transcription factor induces the heat shock operon, increasing heat shock protein levels in the cell. The

stabilization of σ^{32} is thought to occur because the proteases that rapidly degrade it at lower temperatures have a higher affinity for the large numbers of unfolded proteins that are formed at higher temperatures than for σ^{32}, thus increasing its half-life.

The mechanisms of degradation and aspects of the regulation of degradation have been elucidated in the case of individual experimental systems as discussed above. Early studies in prokaryotes have provided observations about changes in general protein stability and postulated the existence of two distinct populations of proteins, i.e. those with a low and those with a high turnover rate (90, 109-111). However, the half-lives of large numbers of proteins in the cell are not known. This is due to the fact that mass spectrometry based proteome-wide techniques to measure half-lives of individual proteins have not been widely applied until very recently (112-121). More knowledge about the half-lives of large numbers of proteins and their regulation in response to different environmental conditions can elucidate how degradation-rates are involved in regulation of protein levels in concert with translational control of (regulated) transcript levels. This in turn will identify proteins for which regulation of degradation-rate plays an important role in their expression level and cellular function.

Measuring synthesis/degradation-rate with pulse-chase Labelling.

From the previous overview of different regulatory points in translation and degradation it is clear that determination of proteome-wide protein synthesis- and degradation-rates can give valuable information on regulation. Together with knowledge regarding changes in protein and transcript levels this will enable not only the identification of proteins that are regulated after transcription but also pinpoint if they are regulated via synthesis, degradation or both.

Determination of protein synthesis-rates usually involves the addition of a labelled compound to the experimental organism of choice in order to distinguish between pre-existing and newly formed proteins. Classically radiolabeled amino acids such as ^{35}S-methionine or a radiolabelled carbon source (i.e. ^{14}C-glucose) are used for this purpose. During the 'pulse', incorporation of radiolabel into proteins can be measured by autoradiography and is a direct measure of protein synthesis provided that pulse times are short (usually 1-2 minutes), to prevent any significant underestimation due to degradation of newly formed proteins (Figure 2a). Conversely, the rate of protein disappearance can be measured following a 'pulse' by the addition of an excess of non-labelled compound preventing further incorporation of radiolabel. During the 'chase' the decline in the radiolabelled protein population is a measure of the rate of protein degradation or protein half-life (Figure 2a). To quantify synthesis and degradation simultaneously a dual labelling method can be used (110), ^{3}H to first label the proteins followed by a change of medium to pulse ^{14}C and chase ^{3}H. This also can be used to determine the relative turnover, irrespective of absolute abundance of the protein by the ratio of ^{14}C/^{3}H or the ratio of newly synthesized proteins over surviving old proteins at a given time point (Figure 2b). This technique, pulse-chase labelling, can be used to measure changes

of the average rate of cellular protein synthesis and degradation by measuring changes in radioactivity of the entire soluble protein fraction of the cell over time. Combination of the approach with immuno-precipitation or protein separation by one- or two-dimensional gel electrophoresis allows determination of protein synthesis and degradation-rates of single or multiple individual proteins (110, 111, 122-125).

Using two-dimensional gel electrophoresis, pulse-chase labelling using radiolabels can also be adopted within a mass spectrometry based proteomics approach (126). The protein synthesis-rate or half-life is determined by autoradiography, followed by excision of Coomassie stained spots on the 2D-gel. The excised spots are digested with trypsin and proteins in the spots identified by mass spectrometry. This approach has several drawbacks though: there are reports that radiolabelling even at low doses used for pulse-chase labelling can generate a stress response, induce cell cycle arrest and even apoptosis (127-132). Furthermore, difficulties to resolve very acidic, basic or hydrophobic proteins (e.g. membrane proteins) on 2D-gels, in addition to protein losses and poor reproducibility limit its quantitative

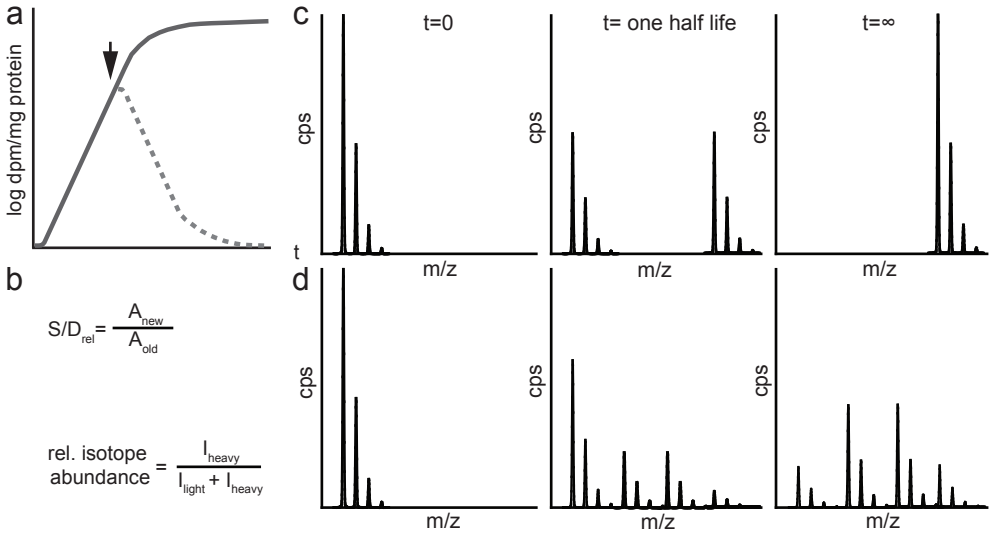


FIGURE 2. **Pulse-chase labelling measures protein synthesis and degradation.** Upon addition of a radiolabel at t=0 a short lag period follows as both unlabeled precursor needs to be depleted and radiolabel needs to be imported into the cell before it can be used in protein synthesis. Subsequently total radioactivity measured within the protein fraction rises as new proteins that are synthesized have radiolabel incorporated (solid line). The radioactivity rises to a plateau when the whole protein population in the cell becomes fully labelled; measuring the rate of increase of radioactivity over a small time interval is a measure of the synthesis-rate. If during labelling an excess of non-labelled compound is added (arrow), then the decline in radioactivity (dotted line) is a measure of the degradation-rate of the proteins **(a)**. The relative rate of turnover can be expressed by the synthesis degradation ratio (S/D_{rel}), expressed as the ratio of abundance of new proteins and old proteins after pulse-labelling. Another way to express the amount of protein turnover after a particular labelling time is the relative isotope abundance, which is the signal intensity of isotopically labelled peptide (I_{heavy}) divided by the total signal intensity **(b)**. Mass isotopomer analysis of the partial labelling pattern obtained for peptides after cells are labelled with **(c)** a heavy amino acid (valine) differing 4 amu from the light counterpart or **(d)** a 1:1 mixed pool of heavy and light valine. In **(d)** there is a 1:3:3:1 ratio of newly synthesized peptides containing 0, 1, 2 or 3 heavy amino acids.

application (133, 134). The possible occurrence of more than one protein in a gel spot, so that the relative contribution of the different proteins to the measured radioactivity cannot be resolved is another intrinsic difficulty. The latter problem could, in extreme cases, lead to incorrect values for synthesis-rates. If a highly abundant protein with a low synthesis-rate would co-migrate with a low abundant protein with a high synthesis-rate, the low abundant protein might well escape mass spectrometric detection, resulting in the attribution of a high synthesis-rate to the abundant protein due to the measured radioactivity in that spot.

Protein turnover measured by stable-isotope incorporation—The use of stable-isotopes in stead of radio-isotopes offers an alternative that is fully compatible with mass spectrometric analysis. Here both identification of proteins and determination of synthesis and degradation-rates come from the mass spectral data acquired. The label can be introduced in various ways, such as a ^{15}N-nitrogen or ^{13}C-carbon source, a stable-isotope labelled amino acid or deuterated water added to growth medium, diet or added intravenously (135, 136). The setup of a stable-isotope labelling experiment is the same as for radiolabelling. The tracer molecule is added at t=0 and samples are taken in time. Protein samples are digested into peptides and subjected to separation before mass spectrometric detection. Database searches with tandem mass spectra of tryptic peptides will lead to identification of proteins. Measurement of the signal intensities of isotopically labelled and non-labelled peptides represent the newly synthesized and surviving old proteins respectively, without the need for double labelling as described for radiolabels.

From the signal intensities of the peptides that represent newly formed and pre-existing proteins the relative turnover (i.e. synthesized and degraded upon a given time after start of the pulse) for the protein identified by the peptides can be calculated (Figure 2b). The synthesis/degradation ratio (115, 120, 121, 136) of newly synthesized proteins over surviving old proteins at a given time point can be expressed analogous to the double radiolabelling technique described above. Relative isotope abundance is expressed by dividing the intensity of the isotopically labelled peptide (newly formed proteins) by the sum of the intensities of the non-labelled peptide (pre-existing proteins) and isotopically labelled peptide (Figure 2b) which results in a turnover relative to the total signal intensity at that time point (116-119). Comparing the relative isotope abundance or synthesis/degradation ratio of cells grown under different conditions reveals changes in protein turnover rate. If labelling times are brief these can be assumed to be relative synthesis-rates (see above). By plotting relative isotope abundances over different time points and correcting for dilution by growth rate of the organism it is possible to calculate the disappearance of pre-existing proteins as well. In this manner a degradation-rate of the protein can be calculated. At steady state (e.g. exponential growth) the half-life found for a protein should be equal to its synthesis-rate (116-119).

There are some advantages to using stable-isotopes instead of radio-isotopes for pulse-chase labelling. First of all no radiation hazard, radiation induced stress response or growth arrest occurs, as alluded to previously. Furthermore, especially in multi-cellular organisms,

the incomplete labelling of the precursor pool for protein synthesis in different tissues can confound analysis of protein synthesis and degradation. This is caused by unlabelled tracer compound that is also being incorporated which can be a problem in both radio and stable-isotope labelling studies. Use of stable-isotopes can circumvent this problem, since it enables determination of the relative isotope abundance in the pool of precursors for protein synthesis in the tissue or cell of interest. The direct precursor pool for protein synthesis are the tRNAs charged with amino acids, but due to their low abundance in the cell measuring their isotopic content is impractical. As such the amino acid pool is often used as a surrogate to measure the relative isotope abundance upon pulse-labelling towards a plateau of full labelling within the experiment (113, 135, 137). However when stable-isotopes are used, as demonstrated for pulse-labelling with an isotopically labelled amino acid in chickens (117), the relative isotope abundance can be ascertained from mass isotopomer distribution analysis (138-140) of partially labelled peptides (Figure 2 c,d). Use of a labelled amino acid is to be preferred over a labelled precursor as an amino acid can be directly used for protein synthesis. While a labelled precursor first has to be metabolized this introduces a significant lag of incorporation that needs to be taken into account. In addition, a labelled amino acid introduces a defined mass shift for all peptides, whereas full ^{15}N or ^{13}C creates different mass shifts for peptides with different elemental composition and a more complex isotope envelope, especially upon partial labelling, complicating analysis.

The use of stable-isotopes as a pulse-label in concert with proteome-wide determinations of the protein turnover rate, using mass spectrometry, has recently been applied in a range of organisms (112-121). As was postulated in earlier studies, two groups of proteins seem to exist (90, 109-111), one large group of proteins with a relatively low turnover rate and a smaller group with a high turnover rate. The added value of the mass spectrometric pulse-chase assay is that the turnover for large numbers of individual proteins under a set of steady state conditions can now be quantified. This is exemplified by the almost 600 proteins for which the half-life was determined in adenocarcinoma cells by Doherty *et al.* (116). Adenocarcinoma cells were fully labelled with ^{13}C-arginine and following an 8 hour chase with ^{12}C-arginine, protein samples taken at various time points were separated by two dimensional gel electrophoresis followed by chromatographic separation coupled to tandem-MS analysis of excised spots. A drawback of the stable-isotope labelling technique is the lengthy labelling time required, as an incorporation of 5-10 percent is required in order to obtain reliable measurements of isotope ratios (135). The detection of small amounts of newly formed proteins in the presence of large amounts of unlabeled proteins is severely limited by the dynamic range of the mass spectrometer (141). This limits the temporal resolution to which stable-isotope labelling can be applied: very short pulse-labelling times are not possible. This means stable-isotope labelling is less equipped to measure transient changes in protein synthesis and degradation following a perturbation, and is more suited to measure overall changes in protein turnover in different steady state growth conditions. Following transient changes in protein synthesis and degradation-rate as an organism adapts

from one state of homeostasis to another, has always been a strong point of radiolabelling. As newly formed and pre-existing proteins can be detected separately, very short labelling times suffice, providing a very high temporal resolution. It is clear from the above that a new approach to pulse-chase labelling that combines the temporal resolution of radiolabelling and the direct compatibility with MS-based proteomics of stable-isotopes can add to the study of posttranscriptional regulation of protein levels.

Non-natural amino acids and detection of new protein formation.

The relative strengths and weaknesses of pulse-chase labelling by stable- and radio-isotopes shows there is a niche for an alternate approach to pulse-chase labelling in the proteomic era. This technique should combine the strong temporal resolution of radiolabelling for the measurement of transient changes in synthesis and degradation-rate with the compatibility of stable-isotope labelling and mass spectrometry based proteomics for the determination of turnover of individual proteins on a truly proteomic scale. The application of such an approach in a wide range of organisms in concert with prior transcriptomic, proteomic and metabolomic approaches would enable searching on a more global scale for different types of post-transcriptional regulation, also in case of transient changes when a cell moves between alternate states of homeostasis. In order to do so, such an approach should combine the direct compatibility with mass spectral detection of the label with the selective detection of labelled species only. The ability to separate labelled species from the bulk of unlabeled material prior to mass spectrometric analysis to increase sensitivity and enable short pulse times would be crucial for such an approach. The label used would have to provide a handle for selective isolation of labelled material. An amino acid as pulse-label has several advantages over labelled precursors of amino acids with respect to incorporation kinetics (see above). With this in mind several required characteristics of such a new pulse-chase label can be formulated. The compound to be used for pulse-labelling is preferentially an amino acid

i) that is efficiently incorporated into proteins
ii) gives minimal disturbance of the structure and function of labelled proteins
iii) differs in mass from its natural counterpart
iv) provides a handle to selectively isolate labelled species
v) can be used in combination with stable-isotopes to quantify protein synthesis-rates between different experiments.

As such a non-natural amino acid seems a natural choice for a pulse-label, as it is both distinct in chemical properties as well as in mass to its natural counterpart. Different non-natural amino acids are known which, when added to the growth medium, can be incorporated into proteins by the translational machinery (142). These non-natural amino acids are ‘close enough analogues’ to be accepted by the aminoacyl-tRNA-synthetase. The

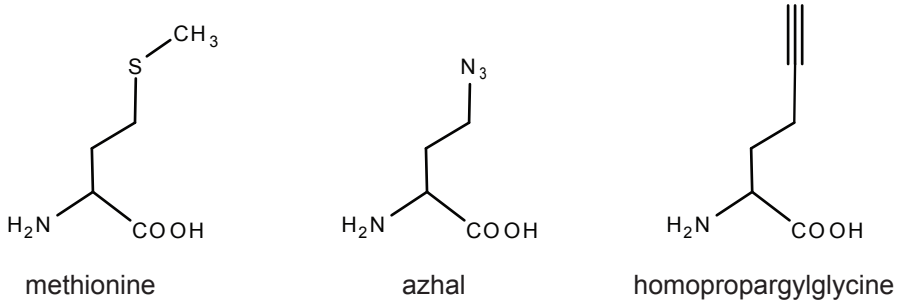


FIGURE 3. **Structure of methionine and two non-natural analogues.**

chemical difference of the non-natural amino acid can offer a handle for isolation of labelled material. By means of a selective chemical reaction directed towards a chemical moiety on the non-natural amino acid, an affinity group of some sort can be introduced to aid in the isolation of labelled proteins or peptides. Both the analogue itself and the specific chemical reaction employed should be bioorthogonal, i.e. show little cross-reactivity with chemical moieties naturally occurring in the cellular environment.

In recent efforts non-natural amino acids have been used to differentiate between newly synthesized and pre-existing proteins. For instance the non-natural amino acid azidohomoalanine (azhal) and homopropargylglycine (HPG), both analogues for methionine (Figure 3), have been reported to be efficiently incorporated into proteins and protein complexes produced in methionine-auxotrophic *E. coli* and mammalian cells grown in the presence of these analogues (143-151). The k_{cat}/K_m of the methionyl-tRNA synthetase is 390 and 500 times lower for azhal and HPG than for methionine in *E. coli* (144). This means auxotrophic organisms are required to ensure no endogenous methionine interferes with labelling of proteins. Both non-natural amino acids contain a group that can be targeted by a specific chemical reaction towards that group. This is an azide group in case of azhal and an alkyne group in case of HPG. The terminal-alkyne group of HPG can react with azides in a copper catalyzed (3+2) cyclo-addition (152, 153). On the other hand, the azide of azhal is amenable to a number of different reactions (Figure 4), namely the before mentioned (3+2) cyclo-addition with terminal alkynes, a ring-strain promoted (3+2) cyclo-addition with cyclo-octynes (154-157) or a Staudinger ligation with phosphines (144, 158-161). Using these different reactions, it is possible to attach a fluorescent group to labelled proteins and in this way visualize new protein formation within the cell or on a cell's surface and also track movement of newly formed proteins over time (148-150, 162-166). These reactions will also be useful in the selective enrichment of newly formed proteins following short pulse-labelling periods by attachment of an affinity handle to labelled proteins or peptides.

Identification of newly synthesized proteins by azhal labelling—Although visualization of newly formed proteins might be useful, for the quantitation of synthesis and degradation-rates of individual proteins mass spectrometric detection is needed. The non-natural amino

acid azhal has been employed in the first mass spectrometry based proteome-wide approaches towards detection of newly formed proteins (141, 167-170). In different studies (see below), it was shown that it is possible to measure newly formed proteins or identify sites of protein turnover in the genome in human endothelial kidney (HEK) cell line (167, 168), cultured *Drosophila melanogaster* cells (171) and *E. coli* (141, 169). Together with reports of incorporation into rat fibroblasts (163) and dissociated hippocampal neuron cultures (166, 167), azhal seems to be suitable for labelling newly synthesized proteins in a variety of biological cell systems. As described above, use of some form of enrichment of labelled proteins is necessary to enable shorter pulse-labelling times and different approaches can be used for this.

Dieterich *et al.* (168) for instance employed enrichment of labelled proteins prior to digestion after pulse-labelling HEK cells for 2 hours with both azhal and $^2H_{10}$-leucine as labels. Following labelling, azhal-containing proteins had an alkyne affinity tag attached (Figure 5a) using copper catalyzed (3+2) cyclo-addition. This affinity tag consisted of an alkyne-reactive group attached via a trypsin cleavable peptide linker to a biotin moiety. Excess reagent was removed prior to loading the tagged protein lysate on an avidin-column, followed by on column tryptic digestion after washing away unbound material. In this manner 195 newly synthesized proteins were identified, by peptides that contained $^2H_{10}$-leucine (28%

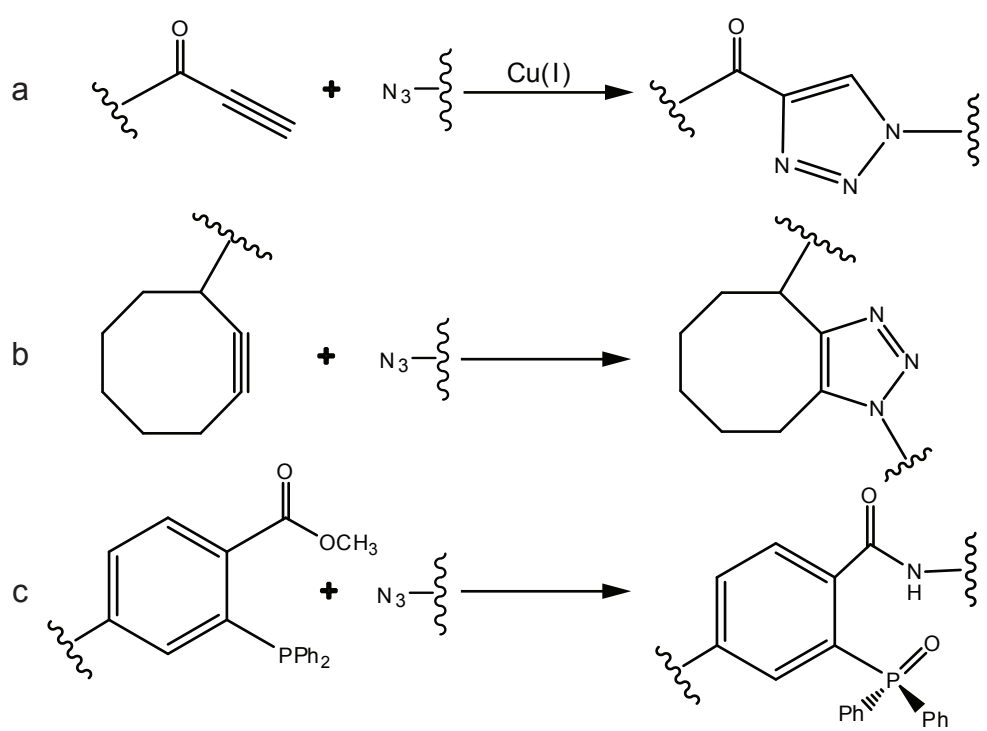


FIGURE 4. **Chemical reactions directed against azhal.** Copper catalyzed (3+2) cyclo-addition between terminal alkyne- and an azide-group (**a**). Strain-promoted (3+2) cyclo-addition between an octyne and an azide (**b**). Staudinger ligation involving an azide reacting with a phosphine (**c**).

of all peptides identified), azhal-containing peptides (3%) or peptides with azhal coupled to remnants of linker (0.08%). The remaining peptides (68.92%) cannot be used to identify newly formed proteins if not found in combination with azhal or $^{2}H_{10}$-leucine containing peptides, as they can also stem from pre-existing proteins. Peptides found containing natural leucine either stem from pre-existing proteins or signify partial labelling due to incomplete labelling of the precursor pool. Using the same enrichment approach with some slight modifications Deal *et al.* (171) used azhal-labelling to identify sites of high turnover of histone proteins in *D. melanogaster.* In this latter study, however, a microarray was used to measure relative abundance of enriched nucleosomal DNA versus total chromosomal DNA on a tiling microarray to identify sites of high turnover of nucleosomal proteins in the genome.

The use of the copper catalyzed (3+2) cyclo-addition seems feasible in combination with azhal-labelling but has drawbacks, such as the reactivity of the Cu^{I} catalyst which can cause side-reactions such as the formation of oxidation products (170). Furthermore we found during the development of a solid phase enrichment approach for azhal-containing molecules that a terminal alkyne attached to a solid phase support via a cleavable trityl-ester linkage (Figure 5b) was prematurely cleaved upon in situ reduction of Cu^{II} to Cu^{I} (172). Taking these early experimental results into account, we further developed a solid-phase enrichment approach which relies on an octyne-group attached to beads via a linker containing a cleavable disulfide bond (Figure 5c). Use of this affinity resin in a peptide-centric rather than a protein-centric approach (170) proved useful in identification of 89 labelled proteins from *E. coli* that grew for one doubling on azhal (50% labelled proteome). Preliminary studies with this resin show promise with regard to reducing pulse-labelling times further towards those common for radiolabelling studies (173).

Outline of this Thesis.

From the work carried out thus far, azhal seems promising for use as a label in a proteome-wide pulse-chase labelling scheme combined with mass spectrometric identification as well as quantitation of synthesis and degradation-rates of individual proteins. However, so far azhal has only been used for proof of principle studies in which newly synthesized proteins were identified, not quantified between cells grown under different conditions, nor was it used to estimate the half-life of large numbers of individual proteins. This thesis deals with the setup and development of a pulse-chase labelling approach with azhal in *E. coli* and most aspects of its application. The effects of growth in the presence of azhal and possible toxic effects of azhal on *E. coli* and *Bacillus subtilis,* two model organisms for Gram-negative and Gram-positive bacteria respectively, are described in *Chapter 2*. The effects of azhal incorporation on the structure and function of four different model proteins was tested as well. *Chapter 3* introduces a specific reaction between tris(2-carboxyethyl)phosphine and azhal-containing peptides and describes three novel reaction products. This reaction forms the basis of an innovative approach to enrichment of azhal-containing peptides. The approach is based on a retention-time shift selectively induced by tris(2-carboxyethyl)phosphine for azhal-containing peptides over two reversed phase chromatographic separations. This approach

was successful in identifying over 500 newly synthesized proteins after only 15 minutes of pulse-labelling with azhal in *E. coli*. In *Chapter 4* this approach is applied to quantitatively measure differences in the levels of newly synthesized proteins after the initial 15 minutes of heat shock in *E. coli* and we compare these data with transcriptomic data from literature. In addition we demonstrate how the approach can be extended by measuring changes in total protein level on the same time-scale to identify stable and labile proteins. The pulse-labelling time is reduced to 10 minutes in *Chapter 5* and the initial response of *E. coli* to a change from an aerobic to an anaerobic environment is described with respect to relative rates of protein synthesis. Stable and labile proteins are identified under these growth conditions and relative synthesis-rates compared to transcript levels upon an anaerobic switch as found in literature. Finally, *Chapter 6* summarizes the advantages and disadvantages for the azhal pulse-labelling approach described here and those of azhal-labelling in general. Furthermore its usefulness as compared to radiolabelling and stable-isotope labelling techniques is discussed, as well as its place in identifying proteins that undergo post-transcriptional regulation. Finally future improvements in its use are suggested as well.

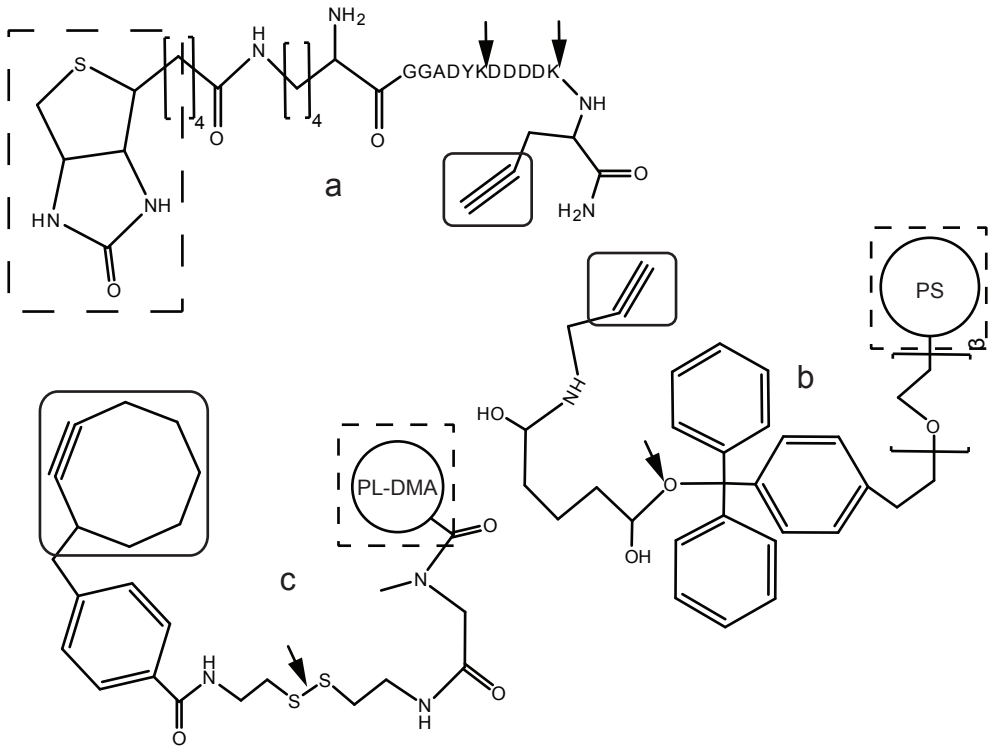


Figure 5. **Affinity resins for the selective enrichment of azhal containing molecules.** Terminal-alkyne biotin tag employed in the BONCAT approach described by Dieterich *et al.* (**a**). Terminal-alkyne coupled to a polystyrene bead (**b**). ARCO-resin employing an octyne reactive group coupled to PL-DMA beads (**c**). Boxes with solid lines: azide-reactive groups, boxes with dotted lines: affinity group or solid-phase resin, arrows denote cleavable sites in the linkers.

2

Effects of azidohomoalanine on bacterial growth, viability and protein function

Summary

The methionine analogue azhal has been used to identify newly synthesized proteins in mammalian cells. The azide group of azhal can be used to specifically enrich labelled proteins, enabling short labelling times. However, not much is known regarding the effects of azhal on cellular physiology and protein function. The suitability of azhal-labelling was tested for two prokaryotic model organisms. Growth in the presence of the analogue was analyzed for methionine-auxotrophic strains of *E. coli* and *B. subtilis*. *E. coli* grows equally well on azhal as on methionine during the first 30 minutes, then gradually growth arrest sets in. Viability and cellular protein content were similar as well, and azhal showed no direct toxicity. In contrast, *B. subtilis* grown on the analogue displayed an initial lag phase and a much lower growth rate. In addition, azhal was toxic at higher concentrations. Nonetheless, *B. subtilis* did also incorporate azhal into proteins.

The effects of azhal on protein structure and function were investigated by producing recombinant proteins in the presence of azhal. Three photo-active proteins: PYP, YtvA and AppA had all their methionine residues replaced by azhal. This shows *E. coli* readily incorporates the analogue into proteins. For all three proteins the UV-VIS spectra were identical to those of their non-labelled counterparts, strongly indicative of correct folding. Recovery from signalling to ground state, after illumination, was slower for azPYP and azYtvA but faster for azAppA compared to the methionine containing proteins. In contrast, recombinant LacZ could not be detected upon induction in the presence of azhal. The initial unperturbed growth on azhal in *E. coli* and the lack of major changes in structure and stability of the three photo-active proteins studied, makes azhal pulse-labelling an excellent tool to determine cellular translation rates.

Introduction

The use of radiolabelled amino acids to determine synthesis and degradation-rates through pulse-chase labelling is a technique which has been used throughout biochemical and physiological research as outlined in *Chapter 1*. Recently however, in efforts to identify and quantitate newly synthesized proteins, the use of non-natural amino acids have emerged as an alternative (141, 167-169). Most notable amongst these non-natural amino acids is the methionine analogue azhal. This methionine analogue was first reported to increase mutational rates in *Salmonella typhimurium* (174), but incorporation into proteins was not studied. Azhal is efficiently incorporated into proteins by both *E. coli*, mammalian and insect cells grown in its presence (141, 143, 144, 148, 169, 171), even though the k_{cat}/K_m of *E. coli* methionyl-tRNA-synthetase for azhal is 390 times lower than for methionine (144). The azide group of azhal makes it amenable to derivatization using different chemistries, facilitating both fluorescent tagging of azhal labelled proteins as well as enhanced mass spectrometric detection of these proteins with different enrichment schemes (see *Chapter 1*). As such, azhal seems a promising label to probe protein synthesis- and degradation-rates on a proteome-wide scale.

However, the suitability of azhal as a pulse-label does not depend solely on its incorporation into (recombinant)-proteins but also on its ability to reflect protein synthesis during pulse-labelling. This is governed by azhal's direct and indirect effects on cellular physiology, through toxicity and relative functionality of azhal-containing proteins. Knowledge about the effects of azhal on cellular physiology is limited, apart from the mutagenicity study in *S. typhimurium* mentioned above, but mammalian cells have been reported to be viable up to two hours after pulse-labelling with azhal as determined by a dye exclusion assay (168). There is somewhat more evidence about how azhal affects the proteins containing it. Display of OmpC at the cell surface of *E. coli* (148, 150) suggests normal folding of azhal-containing proteins as does a study which showed that viral envelope proteins containing azhal assemble into a viral envelope structure (151). A study by Wang *et al.* showed that a recombinant protein containing azhal undergoes N-terminal processing (175) and Dieterich *et al.* did not find large differences in degradation-rates of azhal-containing proteins compared to their unlabelled counterparts in mammalian cells (168). The only report regarding effects of azhal incorporation on specific enzyme activity is on lipase B from *Candida antartica* expressed in an auxotrophic *E. coli*, grown in the presence of azhal, which was found to be 75% that of the methionine containing enzyme (176). Until now azhal has only been applied as a pulse-label in *E. coli,* mammalian and insect cells, (141, 168, 169, 171). So investigating the wider application of azhal in different organisms is important.

Here we report on the potential usefulness of azhal as a pulse-label in two prokaryotic model organisms, i.e. the Gram-negative bacterium *E. coli* and Gram-positive bacterium *B. subtilis*. We measured growth of the bacteria in the presence of azhal to determine the time-frame in which azhal-labelling could give an accurate indication of protein synthesis-rate without interfering with cellular physiology. In addition, effects of incorporation of azhal on protein structure, function and stability were studied with recombinant proteins. We studied three photo-active proteins, i.e. photo-active yellow protein (His-tagged PYP, 15.8 kD, 6 methionines, monomer) from *Halorhodospira halophila* (177), the antirepressor of ppsR, sensor of blue light (AppA His-tagged BLUF-domain, 15.4 kD, 5 methionines, dimer) from *Rhodobacter sphaeroïdes* (178) and the blue-light photoreceptor (His-tagged YtvA, 30.5 kD, 8 methionines, dimer) from *B. subtilis* (179) in addition we tested β-galactosidase (LacZ, 116 kD, 23 methionines, tetramer) from *E. coli*. The photo-active proteins undergo conformational changes upon illumination with light and go through a photocycle (Figure 1) that can be studied by UV-VIS spectroscopy to find possible changes induced by azhal incorporation. For measuring β-galactosidase activity a colorimetric assay was used.

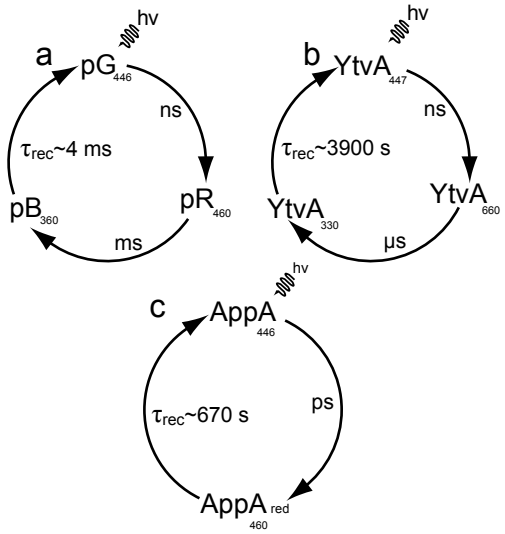


FIGURE 1. **Photocycles of PYP, YtvA and AppA.** Basic photocycle of PYP showing the ground state pG (P or PYP) and key intermediates pR (I_1 or PYP_L) and pB (I_2, PYP_M) formed upon illumination of the protein with photoflash light (**a**). pB is PYP's signalling state in which the chromophore (coumaric acid) has been protonated and the protein structure has changed. The absorbance maxima for the intermediates are shown, the time-scales for the conversion from pG to pR and pR to pB are given, as well as the mean lifetime of the recovery from pB to pG (189) which was measured in the current study to see effects of azhal incorporation on the photocycle. Photocycle of YtvA, showing the ground state ($YtvA_{447}$) and the time-scale upon which intermediates $YtvA_{660}$ and signalling state $YtvA_{330}$ are formed upon illumination (**b**). The mean recovery time from $YtvA_{330}$ to $YtvA_{447}$ is given, numbers denote the absorbance maxima of different intermediates and ground state (209, 210). AppA's photocycle is shown (**c**) with the ground state (AppA) and the red-shifted signalling state ($AppA_{red}$) with their absorbance maxima and mean lifetime of the recovery to the ground state (200, 211, 212).

RESULTS

Growth rate and viability of E. coli cultured on azhal— E. coli has previously been shown to efficiently incorporate azhal into recombinant proteins (143-151). However, for pulse-labelling applications under relevant physiological conditions, it is important to know how *E. coli* grows on azhal and incorporates it into cellular proteins. The K_{cat}/K_m of methionyl-tRNA-synthetase for azhal is 390 times lower than for methionine (144). No residual methionine should thus be present during labelling to ensure maximum efficiency of label incorporation. Therefore a methionine-auxotrophic *E. coli* strain is needed for growth experiments. The methionine auxotroph MTD123 (180) was selected to ensure that no residual methionine was present, cells were harvested and washed prior to addition of azhal as described in *experimental procedures*. A range of azhal-concentrations, from 10 mg/l to 1000 mg/l was tested. Both the growth rate (Figure 2a) and the growth yield (data not shown) are maximal at azhal concentrations of 250 mg/l and higher. For further experiments an azhal concentration of 400 mg/l was chosen.

When *E. coli* cells are growing rapidly, incorporation of radiolabelled compounds into proteins is closely related to growth rate (38). Thus, the degree of azhal-labelling can

be estimated by the increase in cell number as measured by optical density. For the first 30 minutes after inoculation cells grow with similar doubling times for azhal and methionine (Figure 2b). In addition, the amount of cellular protein per OD_{600} is also similar for cells grown on azhal (Figure 3a). After 30 minutes there is a marked decrease in growth rate of the cells growing on azhal. These cells, after having completed more than one doubling at the reduced growth rate, gradually enter stationary phase after more than five hours. Similar results were obtained with the auxotrophic strains CAG18941 and M15MA. This demonstrates that growth rate and increase in protein content in the presence of azhal is similar to that of cells grown on methionine, but only for the first 30 minutes upon incubation of cells with the methionine analogue.

To investigate if cells are still viable after labelling with azhal, samples were taken during growth. Up to 30 minutes, the number of viable cells per unit of OD_{600} remains constant between cells grown on azhal and methionine (Figure 3b). The number of viable cells grown on azhal decreased after one hour of labelling. Toxic effects not related to incorporation of azhal into proteins were tested by growing cells on a mix of azhal and methionine. Because of the lower efficiency of charging azhal to tRNA by the methionyl-tRNA-synthetase,

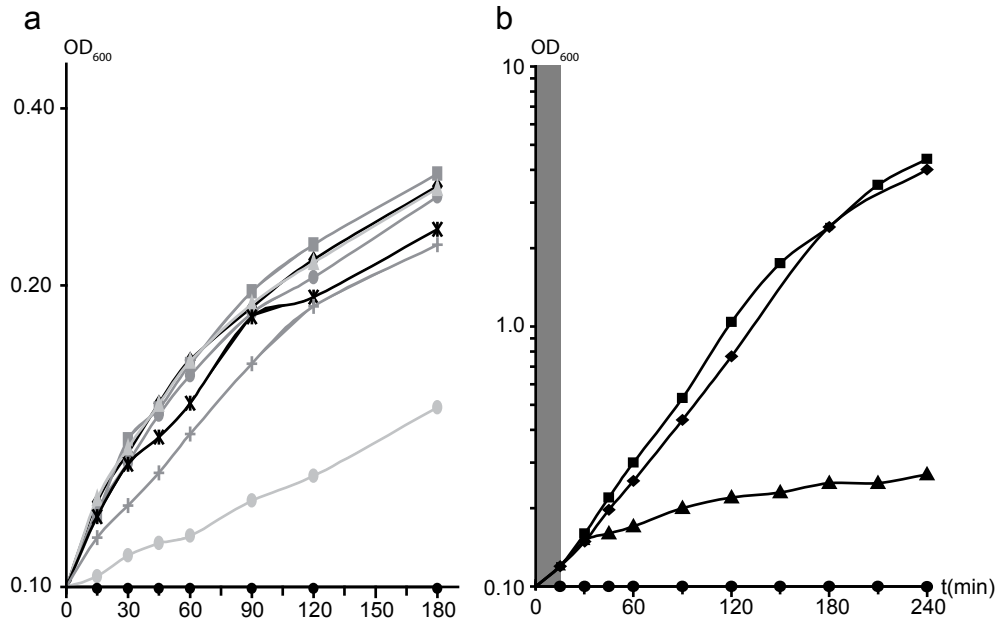


FIGURE 2. **Growth of *E. coli* on different concentrations of azhal**. Growth of *E. coli* strain MTD123 on 10 mg/l azhal (light grey circles), 50 mg/l azhal (dark grey crosses), 100 mg/l azhal (black crosses), 250 mg/l azhal (light grey triangles), 400 mg/l azhal (dark grey squares), 500 mg/l azhal (dark grey circles), 1000 mg/l (black diamonds) or control (black circles) (**a**). No alteration in growth rate is observed above 250 mg/l of Azhal. Growth curve of *E. coli* strain MTD123 on minimal medium containing either 60 mg/l Methionine (squares), 400 mg/l azhal (triangles), a mix of 60 mg/l methionine and 400 mg/l azhal (diamonds) or neither as a negative control (circles) (**b**). Growth rate on methionine alone is 1.15 h^{-1} ± 0.04. Growth rate on azhal plus methionine is 1.11 h^{-1} ± 0.03. The grey box shows the time frame envisioned for pulse-labelling cells with azhal.

incorporation into proteins is negligible under these conditions. Growth rate was compared with cells grown on methionine alone and is similar (Figure 2b). No effects of azhal not incorporated into proteins on cellular physiology are apparent from this experiment.

As an alternative for batch cultures combined with washing to remove excess of methionine prior to azhal-labelling, methionine limited continuous cultures can be used. These enable extremely short pulse-labelling times, since large cultures at high cell densities can be employed, while possible washing-related stresses are avoided. For such a future application we tested strains in methionine limited chemostat culture. We found CAG19491 and M15MA to be unstable auxotrophes in continuous culture. These strains regained the ability to produce methionine endogenously within 24-48 hours of continuous culturing. Although both strains perform well in batch experiments and in producing azhal-labelled recombinant proteins (see below), only MTD123 was used for further labelling experiments.

The work described thus far has been carried out with the MTD123 strain, but growth characteristics of the wild type *E. coli* K12 strain suggests that application of azhal pulse-labelling does not have to be confined to auxotrophes. When grown in minimal medium containing methionine and shifted to minimal medium without methionine, *E. coli* K12 cells show a lag phase of ~15-30 minutes before resuming growth (Figure 4a and b). This is due to the time needed to turn on the methionine bio-synthesis pathway and generate enough endogenous methionine for growth. In contrast, wild type cells shifted to minimal medium containing azhal resume growth immediately at the same rate as the auxotrophic strain for the first 30 minutes (Figure 4b). After 30 minutes the growth of *E. coli* K12 grown on azhal also starts to slow, however, presumably because endogenous methionine biosynthesis is switched on, it recovers growth rate (Figure 4a). This is in accordance with the results obtained for

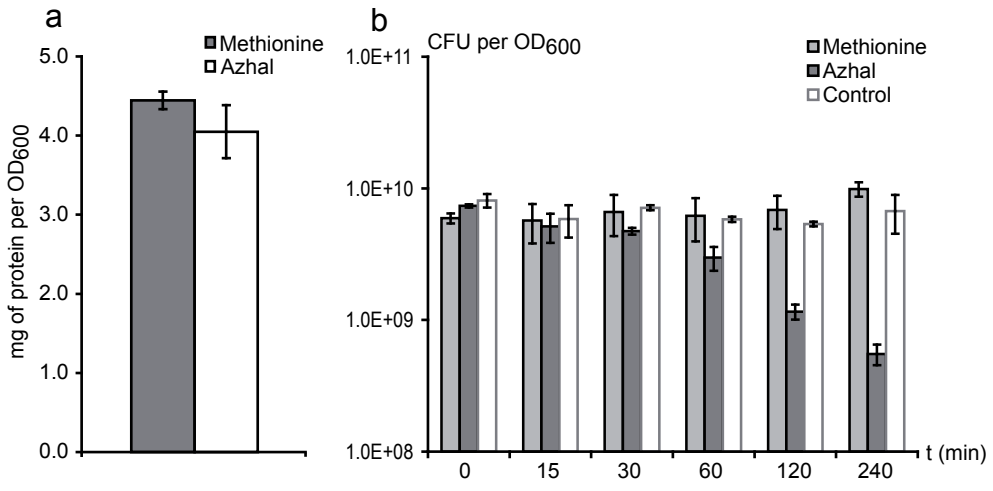


FIGURE 3. **Viability and protein content of *E. coli* grown on azhal**. The amount of total cellular protein per OD_{600} of cells grown on azhal or methionine for the first 30 minutes of growth (**a**). Number of colony forming units per OD_{600} per ml of culture over time. MTD123 cells grown on 60 mg/l methionine or 400 mg/l azhal compared to cells that are in stationary phase in minimal medium without methionine or azhal (**b**).

the auxotroph cultured on a mix of methionine and azhal (Figure 2b). This suggests that wild type *E. coli* strains can be employed as well, provided that labelling times with azhal are short enough to prevent incorporation of methionine in newly synthesized proteins in the course of the pulse. However, we chose to continue the use of the auxotrophic strain to preclude any endogenous methionine biosynthesis interfering with azhal-labelling.

In the experiments described in the previous paragraphs, we have shown that methionine-auxotrophic *E. coli* strains grew at a normal rate for 30 minutes on azhal as a methionine analogue in M9 minimal medium supplemented with all 19 other natural amino acids. These experiments pave the way to use azhal in pulse-labelling experiments in order to identify and quantify proteins synthesized in a short time frame upon an environmental perturbation. Most often stable-isotopes are used for relative quantitation by mass spectrometry. Several methods have been described, either using *in vivo* (181, 182) or *in vitro* (25, 183-186) labelling. *In vivo* labelling has the advantage that pulse-labelling with azhal and introduction of the stable-isotopes occur simultaneously. One possibility for *in vivo* labelling would be application of heavy and light variants of azhal. However, the heavy variant would require expensive ^{13}C- and/or ^{15}N- labelled precursors for its synthesis. Another possibility would be metabolic labelling with stable-isotopes by using a ^{13}C carbon or ^{15}N nitrogen source in the minimal growth medium in one condition and the ^{12}C and ^{14}N counterparts

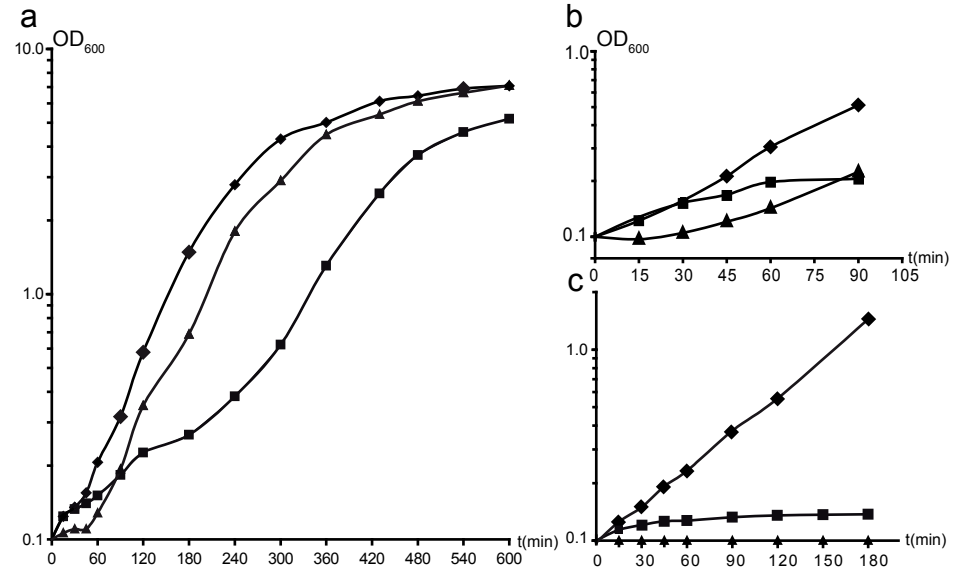


FIGURE 4. **Growth of wild type *E. coli* on azhal and growth of auxotrophic *E. coli* in medium without amino acids in the presence of azhal**. Long term growth of wild type *E. coli* strain K12 on minimal medium containing 60 mg/l methionine (diamonds), 400 mg/l azhal (squares) or neither as a negative control (triangles) (**a**). The initial 90 minutes of the curve shown in (a), shows that growth rate on azhal is similar to growth on methionine for the first 30 minutes, while the negative control resumes growth after a lag phase of about 15-30 minutes (**b**). Growth of the methionine-auxotrophic strain MTD123 on minimal medium not supplemented with 19 amino acids (i.e. all except methionine), containing either methionine, azhal or neither as a negative control (**c**), symbols as in panel (**a**).

in the other. In the experiments with azhal described thus far, *E. coli* was cultured on M9 minimal medium supplemented with 19 amino acids. However, the presence of amino acids in the medium hampers labelling of proteins with stable-isotopes, unless very expensive ^{13}C- or ^{15}N-labelled amino acids are applied. So, to investigate if metabolic labelling could be used to quantify azhal-labelled proteins, MTD123 was grown on M9 minimal medium without addition of the 19 amino acids but containing methionine. Following washing as described in the *experimental procedures* section, cells were transferred to M9 minimal medium containing 400 mg/l azhal without amino acids and OD_{600} was recorded (Figure 4c). Although growth on azhal initially is still similar to growth on methionine, it is only so for 15 minutes following inoculation (Figure 4c). After that, cells cultured on azhal stop growing and enter stationary phase. This shows that addition of amino acids to the medium attenuates azhal-induced growth arrest and indicates that metabolic labelling with ^{13}C and or ^{15}N-labelled amino acid precursors as a quantitative approach is not useful in conjunction with azhal pulse-labelling. Why the addition of amino acids attenuates growth arrest is not entirely clear. Since *in vivo* labelling to introduce heavy and light stable-isotopes turned out to be unpractical in combination with the use of azhal as a pulse-label, we choose application of an *in vitro* method, i.e. the use of iTRAQ (isobaric tagging reagents for relative and absolute quantitation) (183) to quantify newly synthesized proteins (i.e., azhal-labelled proteins) in the experiments described in the following chapters.

B. subtilis grows on azhal and incorporates it into cellular proteins— The Gram-negative model organism, *E. coli,* grows normally on azhal for the initial 30 minutes. Next, Gram-

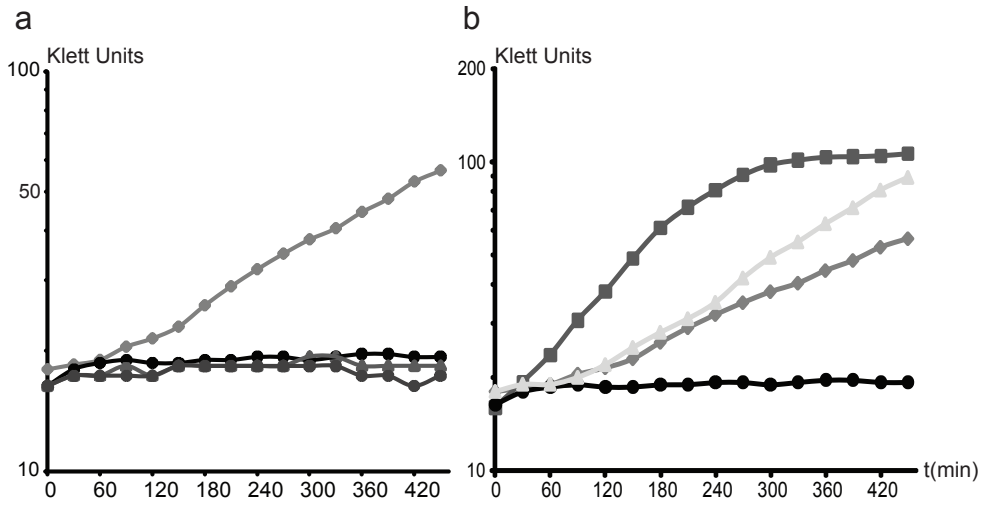


FIGURE 5. **Growth of the methionine-auxotrophic *B. subtilis* strain BR151:*yitJ-lacZ* grown in the presence of azhal**. Growth of *B. subtilis* in the presence of 200 mg/l azhal (grey diamonds), 800 mg/l azhal (grey triangles) or 2000 mg/l azhal (grey circles). Black circles, control (**a**). Growth of *B. subtilis* in minimal medium containing either 50 mg/l methionine (grey squares), 200 mg/l azhal (grey diamonds), a mix of 50 mg/l methionine/200 mg/l azhal (grey triangles) or neither, as a negative control (black circles) (**b**).

positive *B. subtilis* was used to see whether this organism also able to grow on azhal and incorporate it. The methionine-auxotrophic strain BR151 (*yitJ:lacZ)* was cultured in Spizizen minimal medium containing methionine, and after washing was transferred to medium containing azhal. Surprisingly, *B. subtilis* does not grow at all on azhal at concentrations of 0.8 g/l or 2 g/l. (Figure 5a) suggesting a direct dose dependent toxic effect of azhal. Cells grown on 200 mg/l of azhal go into lag-phase for about an hour before resuming growth whereas cells grown on methionine resume growth immediately (Figure 5b, $t_{doubling}$ 1.5 hours). After the lag-phase cells with azhal in the medium resume growth at a lower rate ($t_{doubling}$ 3.75 hours) but continue to grow for over six hours without any sign of going into lag phase. When *B. subtilis* was grown on a mix of azhal and methionine, the initial lag phase remained. Although growth rate did increase ($t_{doubling}$ 2.5 hours) it did clearly not increase to the growth rate of cells grown on methionine alone. This suggests either competition between azhal and methionine for charging to the methionyl-tRNA-synthetase, which is unlikely given the results for growth on different concentrations of azhal, or a direct toxic effect of azhal which is independent of its incorporation into proteins, consistent with its dose dependent cytostatic effect. Nevertheless, growth of this auxotrophic strain on azhal suggests azhal is incorporated into cellular proteins. This indeed proved to be the case (see below).

The *Bacillus* strain used also contained a *yitJ:lacZ* fusion. YitJ is a homocysteine s-methyltransferase and catalyzes one of the steps in the bio-synthesis of methionine in *B. subtilis*. Using the fusion with LacZ, the activation of methionine biosynthesis in response

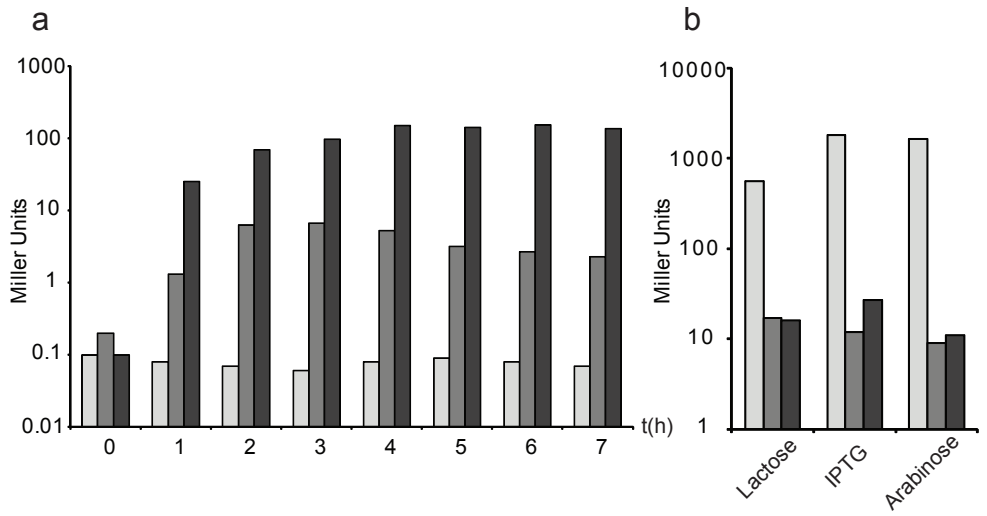


FIGURE 6. **Induction of methionine biosynthesis in *B. subtilis* and** LacZ **activity in *E. coli* grown in the presence of azhal.** LacZ activity as a measure of the induction of the methionine biosynthesis route measured in *B. subtilis* methionine-auxotrophic strain BR151:*yitJ:lacZ* grown in minimal medium containing methionine (light grey bars), azhal (dark grey bars) or neither (black bars) (**a**). LacZ activity measured 3 hours after induction of *E. coli* strain CAG18491/LacZ with lactose, IPTG or arabinose in minimal medium containing methionine (light grey bars), azhal (dark grey bars) or neither (black bars) (**b**).

to methionine starvation can be measured. Figure 6a shows that in a methionine containing medium YitJ expression is repressed, while it is induced in the absence of methionine. Nevertheless cells cannot grow in a methionine-deprived medium due to a mutation in *metB,* encoding a putative homoserine o-acetyltransferase, which blocks the homocysteine biosynthesis pathway. There is also induction of *yitJ* in methionine-deprived cells in the

TABLE I
Azhal containing proteins identified in B. subtilis

No.	Acc. No.	Description	Score†	No. Matches
1	Q04747	Surfactin synthetase subunit 2	1532	35
2	P33166	Elongation factor Tu	1205	22
3	P26901	Vegetative catalase	609	15
4	P27206	Surfactin synthetase subunit 1	527	11
5	Q9KWU4	Pyruvate carboxylase	423	10
6	P28598	60 kDa chaperonin	664	9
7	P02968	Flagellin	465	9
8	P80866	Vegetative protein 296	280	9
9	P37527	Pyridoxal biosynthesis lyase pdxS	358	7
10	O32162	YurU protein	230	7
11	Q08787	Surfactin synthetase subunit 3	453	6
12	P80700	Elongation factor Ts	185	6
13	P49814	Malate dehydrogenase	368	5
14	P21879	Inosine-5'-monophosphate dehydrogenase	216	5
15	P42318	Uncharacterized protein yxjG	212	5
16	P37809	ATP synthase subunit beta	169	5
17	P19669	Transaldolase	139	5
18	P45740	Thiamine biosynthesis protein thiC	268	4
19	P39120	Citrate synthase 2	218	4
20	P09124	Glyceraldehyde-3-phosphate dehydrogenase 1	203	4
21	P37571	Neg. regul. of genetic competence clpC/mecB	162	4
22	P80865	Succinyl-CoA ligase subunit alpha	147	4
23	P40780	Uncharacterized protein ytxH	230	3
24	P21880	Dihydrolipoyl dehydrogenase	206	3
25	P42974	NADH dehydrogenase	139	3
26	O32156	Unchar. ABC transp. Extracell.-bind. protein yur	131	3
27	P39644	Bacilysin biosynthesis oxidoreductase ywfH	122	3
28	P21882	Pyruvate dehydrogenase E1 comp. subunit beta	119	3
29	P80239	Alkyl hydroperoxide reductase subunit C	118	3
30	O31632	YjcJ protein	103	3

†MASCOT protein score based only on azhal-containing peptides, at least 2 peptides per protein

presence of azhal, although to a lesser extent than in the absence of both methionine and azhal. This suggests that cells do sense the lack of methionine and turn on methionine biosynthesis even in the presence of the methionine analogue.

The growth on azhal, after a lag-time of approximately 1 hour, implies that *B. subtilis* incorporates azhal into cellular proteins. This was examined by extracting proteins from cells grown for 8 hours on the analogue and digesting the extracted proteins with trypsin. Following tryptic digestion azhal-containing peptides were enriched and analyzed by mass spectrometry as described in detail in *Chapter 3*. Following isolation and mass spectrometric analysis, 626 unique peptides were identified of which 223 were found to contain azhal. These azhal-containing peptides were derived from 103 different proteins of which 54 were identified by two or more azhal-containing peptides (Table I).

TABLE I
continued

No.	Acc. No.	Description	Score†	No. Matches
31	P39126	Isocitrate dehydrogenase [NADP]	101	3
32	P09339	Aconitate hydratase	138	2
33	O31629	UPF0477 protein yjcG	129	2
34	P80868	Elongation factor G	118	2
35	P31104	Chorismate synthase	113	2
36	P13243	Probable fructose-bisphosphate aldolase	111	2
37	P04969	30S ribosomal protein S11	104	2
38	P49786	Biotin carboxyl car. prot. of ac.-CoA carboxylase	104	2
39	P02394	50S ribosomal protein L7/L12	101	2
40	O34660	Aldehyde dehydrogenase	96	2
41	P55873	50S ribosomal protein L20	89	2
42	P37808	ATP synthase subunit alpha	85	2
43	P06224	RNA polymerase sigma factor rpoD	82	2
44	O32157	YurP protein	82	2
45	P21471	30S ribosomal protein S10	81	2
46	P28015	Putative septation protein spoVG	79	2
47	O34934	Prob. inorg. polyphosphate/ATP-NAD kinase 2	77	2
48	O32174	Glycine cleavage system H protein	75	2
49	P08838	Phosphoenolpyruvate-protein phosphotransferase	75	2
50	P80643	Acyl carrier protein	69	2
51	P18255	Threonyl-tRNA synthetase 1	68	2
52	P80698	Trigger factor	65	2
53	Q06797	50S ribosomal protein L1	64	2
54	P80859	6-phosphogluconate dehydrog, decarboxylating 2	58	2

†MASCOT protein score based only on azhal-containing peptides, at least 2 peptides per protein

Incorporation of azhal into LacZ and effects on protein function— B. subtilis incorporates azhal into proteins as shown above. Indeed, the LacZ expression of cells grown on azhal suggested that methionine biosynthesis was not induced to the same extent as under full methionine starvation conditions without azhal. However, for the LacZ activity measurements of *B. subtilis* BR151 grown with azhal to be a reliable indication of *yitJ* expression, enzymatic activity of LacZ containing azhal should be comparable to activity of LacZ containing methionine. As the fusion gene used in *B. subtilis* stems from *E. coli*, the induction of the lac-operon in *E. coli* was tested in the presence of azhal. Cells grown on M9 minimal medium containing 19 amino acids were shifted to lactose-containing medium containing either methionine, azhal or neither as a negative control for three hours. Subsequently cells were

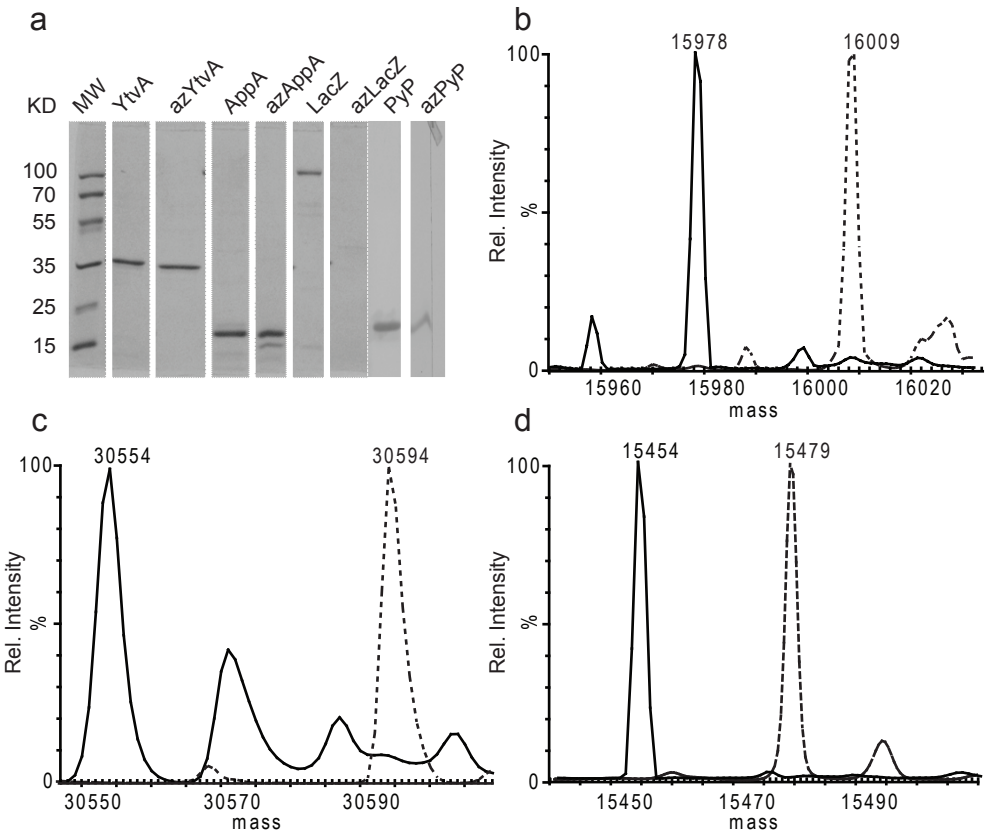


FIGURE 7. **Incorporation of azhal into recombinant proteins produced in *E. coli*.** SDS-PAGE gel after Ni-column purification of His-tagged recombinant proteins produced in *E. coli* grown in minimal medium containing either methionine or azhal (**a**). Overlay of deconvoluted-mass spectra of PYP (dotted line) and azPYP (solid line), theoretical mass of His-tagged PYP and its chromophore *p*-coumaric acid attached is 16007 Da (**b**). The mass difference of -31 amu shows that all six methionine residues have been replaced by azhal (mass difference azhal and methionine is -5 amu). Overlay of deconvoluted-mass spectra of YtvA (theoretical mass: 30593 Da) and azYtvA (**c**). The mass difference of -40 amu shows all eight methionine residues were replaced. Overlay of deconvoluted-mass spectra of $AppA_{5\text{-}125}$ (theoretical mass: 15479 Da) and $azAppA_{5\text{-}125}$ (**d**). The mass difference of -25 amu indicates that all 5 methionine residues have been replaced by azhal.

disrupted and assayed for LacZ activity. Clearly, cells in which the lac-operon is induced in the presence of azhal have LacZ activity comparable to that of the negative control, whereas cells grown in methionine containing medium show high LacZ activity (Figure 6b). To exclude that lactose import through the lactase transporter (LacY) prevents full induction of the operon, e.g. azhal-containing LacY may not import lactose efficiently, cells were also induced by the lactose analogue isopropyl β-D-1-thiogalactopyranoside (IPTG), not requiring a transporter for membrane passage (Figure 6b). Again cells grown on azhal did not show more induction of LacZ-activity than the negative control upon induction by IPTG. These results indicate that LacZ containing azhal has no, or a severely reduced, enzymatic activity.

To test whether azhal incorporation into LacZ leads to a markedly less functional protein we transformed *E. coli* with a pBAD plasmid containing LacZ under an arabinose-inducible promotor and a His-tag which enables purification for in vitro assays. Induction of LacZ production with arabinose again showed that cells grown in azhal-supplemented medium had the same level of LacZ activity as the negative control, whereas cells grown on methionine containing medium showed high levels of LacZ activity (Figure 6b). However, when LacZ was purified from the soluble protein fraction by Ni-affinity columns, no azhal-containing LacZ was recovered whereas large amounts of methionine containing LacZ could be purified (Figure 7a). This suggests that azhal-containing LacZ probably does not assemble correctly and precipitates into inclusion bodies and/or is rapidly degraded. Therefore, LacZ is not a suitable reporter to measure azhal induced activation of methionine biosynthesis in *B. subtilis.*

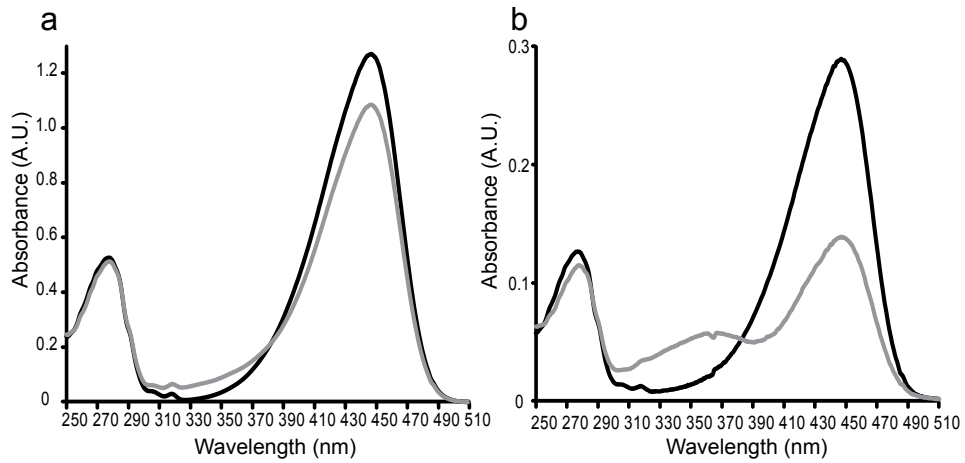


FIGURE 8. **UV-VIS Spectra of PYP and azPYP.** Spectrum of ground state (black line) and light activated state (grey line) after photoflash excitation of PYP (**a**) or azPYP (**b**). Ground state spectra of both PYP and azPYP are highly similar with absorbance maxima at 280 and 446 nm. Both have a marked reduction in absorbance at 446 nm and a blue shift to ~355 nm upon photoflash excitation, showing azPYP to be photo-active, with the same iso-bestic point at 383 nm.

Incorporation of azhal into photo-active proteins, effects on spectral properties and photocycle kinetics— To further study the effects of azhal incorporation on protein structure and function, we also produced three photo-active proteins in *E. coli* in the presence of azhal, under an IPTG inducible promotor. The three photo-active proteins are PYP (*H. halophila*), YtvA (*B. subtilis*) and the N-terminal BLUF domain of AppA (*Rb. sphaeroïdes*). Induction of expression in azhal-containing medium followed by reconstitution with their respective chromophores and Ni-column purification yielded three proteins (Figure 7a). Measured by their average mass shift compared to their methionine containing counterparts these proteins had all their methionine residues replaced by azhal (Figure 7b, c, d). This clearly demonstrates that methionine-auxotrophic *E. coli* incorporate azhal into proteins when grown in the presence of the methionine-analogue, and that these azhal-containing proteins are not all rapidly degraded or precipitate as seems to occur with LacZ.

To ascertain whether changing the methionine residues to azhal influences azPYP's structure around its chromophore, *p*-coumaric acid, the UV-VIS spectra of PYP and azPYP are compared in Figure 8. As is apparent from the ground state spectra, absorption maxima have not changed. Moreover, both proteins are photo-active and show bleaching at 446 nm and a blue shift to ~355 nm (pB) after excitation of the protein solution by photoflash illumination (Figure 1a). Subsequently, the recovery to the ground state was measured by determining the rate of recovery of absorbance at 446 nm upon photoflash excitation. Figure 9 and Table II show that the rate of recovery of azPYP (0.89 s^{-1}) is slightly lower than the rate of recovery measured for methionine containing PYP (1.28 s^{-1}). The rate measured for methionine containing PYP is somewhat less than the 2.0 s^{-1} to 3.4 s^{-1} reported in literature

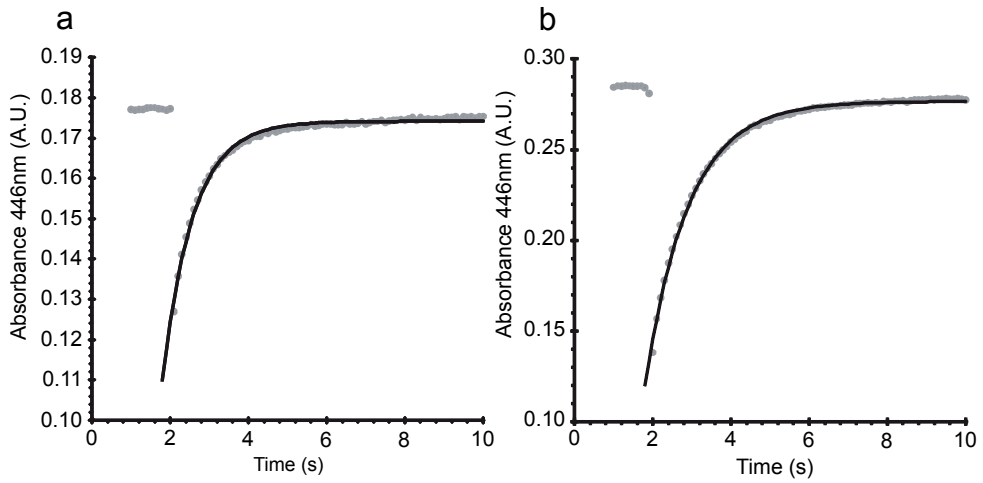


FIGURE 9. **Recovery of the absorbance at 446 nm of PYP and azPYP.** Figure shows the recovery of absorbance at 446 nm after photoflash excitation of PYP (**a**) and azPYP (**b**), grey dots are absorbance values measured before and after the excitation of the protein (at the 2 second time point), the black line is the fit of the mono-exponential function by Excel least squares analysis to the absorbance data. The recovery rates of these fits are presented in Table II.

(187-189). This could be due to differences in the purification procedure. Proteins in this study were only enriched on a Ni-column, yielding two variants with respect to the rate of recovery to the ground state (M.A. van der Horst, personal communication). Further purification can separate the slower recovering variant from the fast recovering one, resulting in a higher overall recovery rate. Furthermore proteins in this study did not have their histidine-tags removed, which also influences recovery rate. The methionine residue at position 100 plays an important role PYP photocycle kinetics through interaction with Arg_{52}, acting in stabilizing the ground state and facilitating the recovery from pB to the ground state after photo-activation (190, 191). Thus, as expected, mutants of M_{100} show a significantly slower recovery to ground state varying from $6.5*10^{-2}$ s^{-1} ($M_{100}E$) to $1.9*10^{-3}$ s^{-1} ($M_{100}L$ and $M_{100}A$) and $1.1*10^{-3}$ s^{-1} ($M_{100}K$) (190, 191). In comparison, replacement of all methionine residues including Met_{100} by azhal only marginally decreases the recovery rate to ground state.

TABLE II
Rate of recovery to ground-state after illumination of photo-active proteins

Protein	$\tau_{recovery}$ (s)	k (s^{-1})	$k_{rel.}$†	λ(nm)
PYP	0.78	1.28	1	446
azPYP	1.12	8.90×10^{-1}	0.69	446
AppA	600	1.67×10^{-3}	1	495
azAppA	175	5.73×10^{-3}	3.44	495
YtvA	3783	2.64×10^{-4}	1	450
azYtvA	4943	2.02×10^{-4}	0.77	450

† recovery rate relative to rate of methionine containing photo-active protein.

The chromophore (coumaric acid) is thought to be in anionic form in the active site and to be protonated during formation of the pB intermediate of the photocycle (192). The protonation state of the chromophore can be influenced by the pH. Lowering the pH results in reversible formation of a blue shifted intermediate at low pH called pB_{dark} (177, 192). Protonation of the chromophore is a cooperative process, presumably because of the existence of an extensive hydrogen-bonding network in the pG state which has to be disrupted (193). The formation of pB_{dark} was measured for azPYP by pH titration (Figure 10a) and the pK_a and Hill coefficient (*n*), which expresses the degree of cooperativity of the protonation were obtained (Figure 10b) and compared to PYP. The Hill coeffient (*n*) of azPYP (*n* =1.8) was the same as for PYP (*n* =1.8) which was consistent with earlier observations (192, 193) of methionine containing PYP. Azhal does not seem to influence the cooperativity of protonation

of the chromophore. However the pK_a of azPYP, (3.2) was slightly higher than the pK_a (2.8) for methionine containing PYP (177, 192, 193). As the pB and pB_{dark} intermediates result from partial unfolding of the protein (192), the slightly higher pKa of azPYP could indicate a somewhat less stable protein structure in azPYP compared to wild type PYP. All in all, azPYP appears to be functionally highly similar to wild type PYP with only slight alterations in photocycle recovery rate and protein stability

The effects of azhal incorporation on spectral properties of YtvA are shown in Figure 11. As for PYP, the ground state spectrum of azYtvA is highly similar to that of methionine containing YtvA. This shows that no major changes in structure occur that interfere with chromophore binding. After illumination, the azYtvA spectrum shifts to the same single maximum absorption at 386 nm as YtvA, confirming that azYtvA is photo-active too. The kinetics of recovery from illuminated to dark state of azYtvA are shown in Table II. The recovery rate of azYtvA ($2.02*10^{-4}$ s^{-1}) is only marginally slower than that of YtvA ($2.64*10^{-4}$ s^{-1}), which shows that replacing methionine residues with azhal does not have a large effect on YtvA's photocycle kinetics.

The spectrum of the BLUF-domain of AppA ($AppA_{5-125}$) produced in *E. coli* grown in azhal-containing medium also is very similar to its methionine containing counterpart (Figure 12). $azAppA_{5-125}$ was photo-active and showed the same ~10 nm red shift upon illumination. However, after illumination the rate of $azAppA_{5-125}$ recovery to the ground state was 3.4 fold faster than that of wild type $AppA_{5-125}$ (Table II). This surprising increase in recovery rate is reminiscent of the increase in recovery rate found for AppA $W_{104}F$ which is 2.7 fold that of wild type (194). Although $azAppA_{5-125}$ does not show a change in the red-shift after illumination, which is the case for the $W_{104}F$ mutant, the increased recovery could also indicate a destabilization of the signalling state (194) by azhal incorporation. In the BLUF-

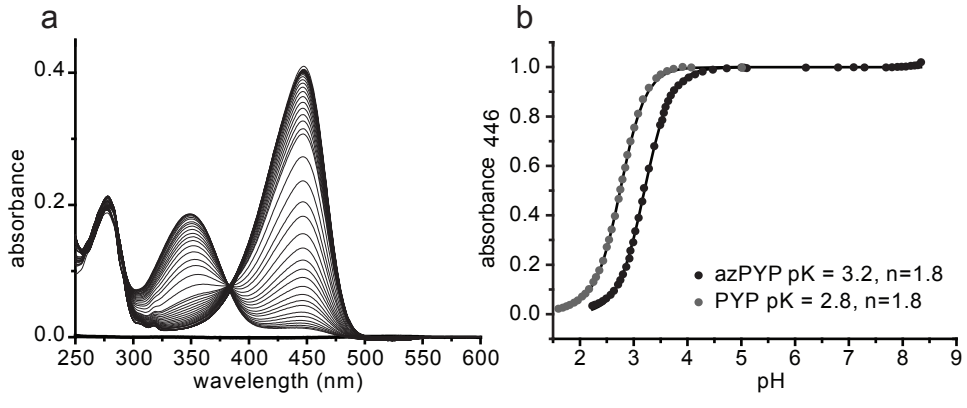


FIGURE 10. **pH titration of the absorption spectra of PYP and azPYP.** Dependence of the absorption spectra of azPYP on pH (**a**). Spectra were taken at room temperature between pH 1 and pH 8.5 at room temperature in 10 mM tris-Cl, 100 mM KCl. The relative amplitude of the absorbance at 446 nm as a function of pH for PYP (grey dots) and azPYP (black dots) (**b**). Theoretical curves (solid lines) were obtained by fitting the data to the modified Henderson Hasselbalch equation [1].

domain protein AppA the methionine at position 106, conserved in other BLUF-domain proteins, is thought to change its interaction with Glu_{63} during light induced conformational changes in the protein (195-197). Although all methionine residues were replaced by azhal, the altered kinetics of the rate of recovery to the ground state does support the notion of the involvement of this specific methionine residue in the conformational changes of the protein in ways that cannot be exactly mimicked by the methionine analogue azhal.

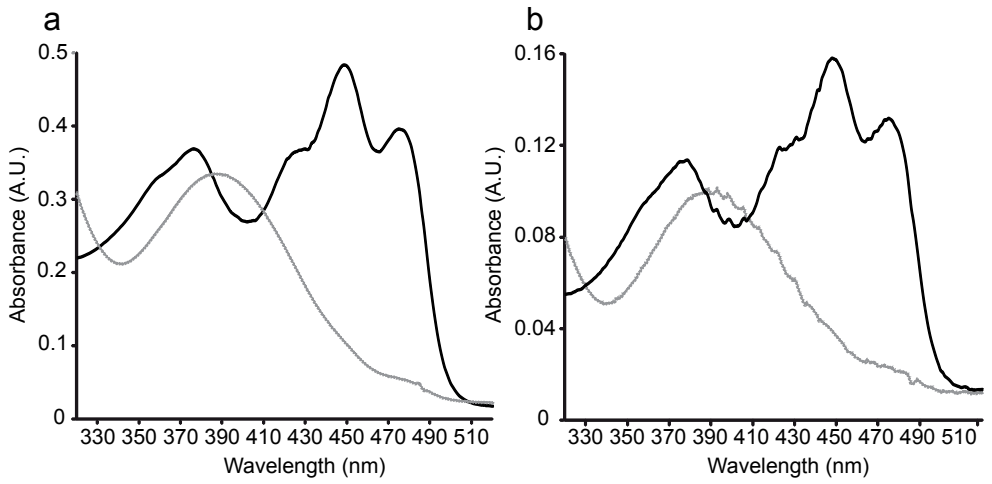


Figure 11. **UV-VIS Spectra of YtvA and azYtvA.** Spectrum of dark state (black line) and illuminated state (grey line) after illumination of YtvA (**a**) and azYtvA (**b**). Ground-state spectra of YtvA and azYtvA are similar with maxima of absorbance at 376, 448 and 475 nm. Just like YtvA, azYtvA shows photo-activity and the illuminated state shows the same single maximum at 386 nm, with iso-bestic points observed at 331, 386 and 409 nm.

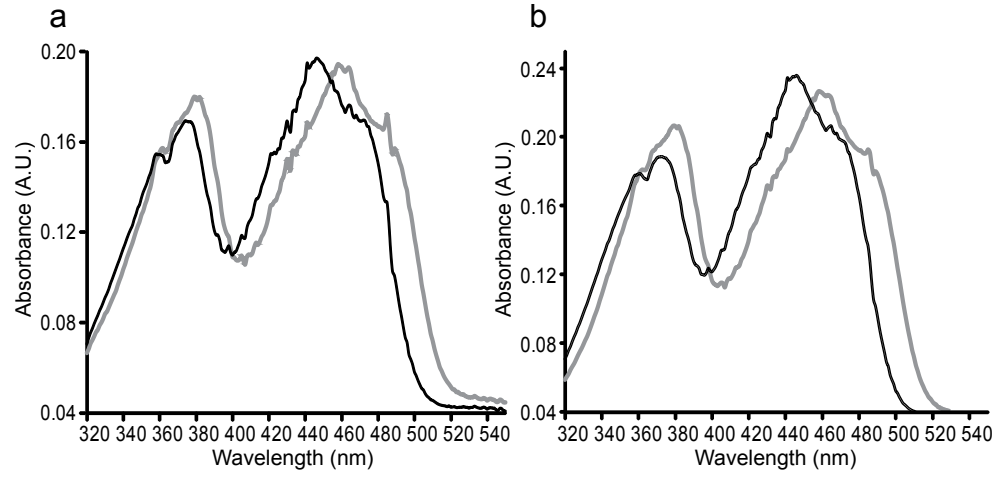


Figure 12. **UV-VIS Spectra of $AppA_{5-125}$ and $azAppA_{5-125}$.** Spectrum of dark state (black line) and illuminated state (grey line) after illumination of $AppA_{5-125}$ (**a**) and $azAppA_{5-125}$ (**b**). Ground-state spectra of $AppA_{5-125}$ and $azAppA_{5-125}$ are very similar, exhibiting major absorbance maxima at 374 and 446 nm. $azAppA_{5-125}$ is photo-active and also shows a red shift ~10 nm to an illuminated state with maxima at 380 and 460 nm, like $AppA_{5-125}$. Both proteins exhibit iso-bestic points at 359, 398 and 454 nm.

Discussion

Azhal's suitability for pulse-labelling was tested by measuring growth and viability of *E. coli* grown on a medium with azhal substituting for methionine. Growth rate, viability and cellular protein content were comparable for cells grown on azhal or methionine for the initial 30 minutes. Free azhal, not incorporated into proteins, gave no evidence of toxicity, since an excess of azhal in the presence of saturating amounts of methionine has no effect on growth rate. After prolonged labelling however, growth arrest does occur, which can be explained by dysfunction of essential proteins with crucial roles for one or more methionine residues. Another possible explanation for the occurrence of growth arrest is the role of methionine as a methyl donor for methylation reactions in the form of s-adenosyl methionine. As is obvious from its structure azhal cannot replace methionine in this function and as such would cause methylation signalling to go awry, possibly causing growth arrest. Nevertheless, because initial growth rate and protein formation seem unaffected, azhal seems a promising label for short time-scale proteome-wide pulse-labelling studies in *E. coli*.

In stark contrast to *E. coli,* growth of *B. subtilis* on azhal at 200 mg/l showed a prolonged lag phase, and growth was less than 3 fold slower. In addition *B. subtilis* did not go into stationary phase in the time-scale measured. Concentrations of 800 mg/l did induce complete growth arrest. Strikingly, no effect on initial growth rate of *E. coli* was observed at this azhal concentration both in the absence and presence of methionine. With *B. subtilis,* mixtures of methionine and azhal gave an initial lag phase and diminished growth rate. This suggests a secondary toxic effect independent of incorporation into cellular proteins, consistent with the absence of growth at higher azhal concentrations. In spite of the observation that *B. subtilis* incorporates azhal into cellular protein, both the different growth on azhal, the suspected secondary toxicity and in particular the initial lag phase, lead to the conclusion that azhal is not the most suitable label for pulse-labelling experiments in *B. subtilis*. This result is surprising as in addition to *E. coli*, growth on azhal has been reported for various eukaryotes (141, 143, 144, 148, 169, 171). In contrast however, we previously found that *Saccharomyces cerevisiae* also showed growth inhibition that was dependent on the concentration of azhal in the medium (Supplemental Figure 1). This shows that although azhal-labelling seems applicable to a variety of organisms, labelling and growth needs to be tested for each prospective organism.

Suitability of an amino acid analogue for pulse-labelling requires that incorporation in proteins should have no important effect on protein folding and stability. The highly different physiological responses of two bacteria growing on the methionine analogue azhal, sparks the question about how incorporation of the analogue affects protein structure and function. This was tested by production of various recombinant proteins in *E. coli* grown in the presence of azhal. In line with earlier reports (143, 144, 148, 176) we observed that *E. coli* readily incorporated the analogue into PYP, YtvA and AppA. Judged from their ground-state UV-VIS spectra these proteins folded normally and retained photo-activity upon illumination. Measurements on the rate of recovery from signalling to ground state for these

proteins showed a slight decrease in recovery rate for azPYP and azYtvA, i.e. these proteins remain in the signalling state somewhat longer. In contrast, the recovery rate of azAppA increased significantly, suggesting strong destabilization of the signalling state. pH-titration experiments with azPYP also suggested a slightly destabilized ground-state structure. However, on the whole the three proteins kept their activity and were correctly folded as observed before with other proteins and protein complexes (144, 148, 151, 176). In sharp contrast to these findings, LacZ activity was not measurable when induced in cells grown in azhal-containing medium and His-tagged azLacZ could not be purified from the soluble protein fraction of *E. coli*. This strongly suggest misfolding and precipitation or accelerated degradation of azhal-containing LacZ.

Though similar in electron density to methionine (144), azhal cannot substitute methionine for all its physio-chemical characteristics in a cellular context as illustrated by the non-functional azLacZ protein. The misfolding and subsequent degradation of a subset of proteins is a possible cause of the eventual growth arrest in *E. coli* after 30 minutes and in proteome-wide pulse experiments these proteins will probably escape detection. However, both the initial growth rate of *E. coli* on azhal and minor effects of azhal incorporation on the structure and function of proteins described here and in literature (148, 150, 151, 176) suggest that it can give an accurate measure of cellular translation rates for most proteins.

Supplemental Data—Supplemental figures and tables can be found in the addendum section on page 134.

Experimental Procedures

Synthesis of L-azhal— L-azhal was synthesized from L-Boc-2,4-di-aminobutyric acid (L-Boc-DAB, Chem-Impex, Wood Dale, USA) by diazotransfer (198) using Triflic azide (TfN_3) as previously described (141).

Strains and plasmids— Chemically competent CAG18491 cells (*E. coli* genetic stock centre Yale, USA) were transformed with pREP4 (Qiagen, Venlo, the Netherlands) to make CAG18491/pREP4. Subsequently, chemically competent CAG18491/pREP4 were transformed with plasmids pHISP (199), pQE30X$_a$/AppA$_{5\text{-}125}$ (200), pQE30X$_a$/YtvA (201) or pBAD/His/LacZ (Invitrogen, Breda, the Netherlands) to generate the protein production strains (Table III).

Cell culture, growth curves, viability- and β-galactosidase assays— *E. coli* strain MTD123 (180) was grown aerobically at 37 °C in LB medium. For growth experiments cells grown overnight in LB medium were transferred to M9 minimal medium containing 6.8 µM $CaCl_2$, 1.0 mM $MgSO_4$, 59.3 µM thiamine·HCl, 57.0 nM Na_2SeO_3, 5.0 µM $CuCl_2$, 10.0 µM CoCl, 5.2 µM H_3BO_3, 99.9 µM $FeCl_3$, 50.5 µM $MnCl_2$, 25.3 µM ZnO, 0.08 µM Na_4MoO_4, 111 mM glucose with 40 mg/l of tyrosine and 60 mg/l of each of the other 19 natural amino acids (Sigma-Aldrich, St Louis, USA). Cells were inoculated at OD_{600} 0.1 and allowed to grow into exponential phase before being harvested at OD_{600} 1.0, by spinning down the cells for 10 minutes at 4500 rpm and 4 °C. Cells were then washed twice by resuspending the cell pellet in sterile M9 medium without additives (followed by centrifugation) to eliminate traces of methionine. After washing, cells were transferred to M9 minimal medium (see above) in which the methionine was replaced by azhal and cells were allowed to resume growth aerobically at 37 °C. Growth curves were recorded with varying methionine or azhal conditions as indicated.

To determine the number of viable cells, cells were diluted in sterile M9 medium and plated on LB-

agar plates in duplo; colonies were counted after overnight growth at 37 °C. β-Galactosidase assays were performed essentially as described by Miller (202). In short, cells (CAG18491/LacZ) were induced with lactose (2% final concentration), IPTG (200 µM final concentration) or arabinose (1% final concentration) for 3 hours. Next, cells were disrupted by the addition of 0.1% SDS, 1% chloroform and placed on ice. Subsequently 1 mM ortho-nitrophenyl-β-galactoside was added and the reaction was allowed to proceed at 37 °C until being stopped by the addition of 1 M Na_2CO_3 to the reactions. B-galactosidase activity was measured at 420 nm and corrected for reaction volume, OD_{600} and reaction time to yield Miller units.

The *B. subtilis* strain BR151 (auxotrophic for met, arg, lys) was grown aerobically at 37 °C in Spizizen minimal medium (203) with addition of 50 mg/l methionine, arginine and lysine. Cells were allowed to grow into early exponential phase before being harvested, and washed twice with minimal medium at room temperature. After washing, cells were transferred to minimal medium (see above) in which methionine was replaced by azhal and cells were allowed to resume growth aerobically at 37 °C. Growth curves were recorded using a Klett-Summerson photoelectric colorimeter with methionine or azhal conditions as indicated. B-galactosidase assays were performed as described above.

TABLE III
E. coli and B. subtilis strains and plasmids used in this study

name	organism	genotype	plasmids	resistance	source
MTD123	*E. coli*	ΔyagD, ΔmetE, ΔmetH	-	-	(180)
CAG18491	*E. coli*	λ^-, rph-1, metEo-3079::Tn10	-	tet	(206, 207)
CAG18491/pRep4	*E. coli*	λ^-, rph-1, metEo-3079::Tn10	pREP4	kan, tet	this study
CAG18491/PYP	*E. coli*	λ^-, rph-1, metEo-3079::Tn10	pREP4, pHISP	amp, kan, tet	this study
CAG18491/APPA	*E. coli*	λ^-, rph-1, metEo-3079::Tn10	pREP4, pQE30X$_a$/APPA$_{5-125}$	amp, kan, tet	this study
CAG18491/YTVA	*E. coli*	λ^-, rph-1, metEo-3079::Tn10	pREP4, pQE30X$_a$/YTVA	amp, kan, tet	this study
CAG18491/LacZ	*E. coli*	λ^-, rph-1, metEo-3079::Tn10	pREP4, pBAD/His/LacZ	amp, kan, tet	this study
M15MA	*E. coli*	Na^{IS}, Str^S, Rif^S, Thi$^-$, Lac$^-$, Ara$^+$, Gal$^+$, Mtl$^-$, F$^-$, RecA$^+$, Uvr$^+$, Lon$^+$. metEo	pREP4	kan	(148)
BR151:*yitJ-lacZ*	*B. subtilis*	trpC2, metB10, lys-3, *yitJ-lacZ*	-	-	(208)

Production and purification of recombinant proteins containing azhal— Cells were grown into early exponential phase at 37 °C in 500 ml M9 minimal medium as described above, adding 60 mg/l methionine, 50 µg/l tetracycline, 50 µg/l ampicillin and 50 µg/l kanamycin. For production of YtvA, cells were grown in the dark and the protein was protected from light throughout the purification procedure and before subsequent measurements. Cells were induced directly in the methionine containing medium with 100 µm IPTG (PYP, AppA, YtvA) or 1% arabinose (LacZ) before a further 3 hours of incubation in a shaking incubator set at 37 °C. Alternatively, cells were harvested, washed twice with M9 buffer, inoculated in 500 ml minimal medium with antibiotics but lacking methionine and incubated at 37 °C for half an hour before addition of 400 mg/l azhal prior to induction. Cells were subsequently harvested

by centrifugation, resuspended in 100 mM tris-Cl pH 8.0, 1 mg/ml lysozyme, 25µg/ml DNase/RNase and lysed by sonication. Lysates were centrifuged at 15000 rpm for 45 minutes at 4 °C to remove cellular debris. Subsequently, photo-active proteins were reconstituted with their respective chromophores as described below before they were loaded on a Ni-agarose column (Qiagen, Venlo, the Netherlands) by gravity flow. Columns were washed with 50 ml 100 mM tris-Cl pH 8.0 before samples were eluted with 8 ml 500 mM imidazole in 100 mM tris-Cl pH 8.0 and dialyzed immediately against 100 mM tris-Cl pH 8.0 in 5kD cut-off dialysis tubing (Spectrum Laboratories, Rancho Dominguez, USA) overnight at 4 °C.

Reconstitution of chromophores in photoproteins— Prior to purification, photo-active proteins were reconstituted with their respective chromophores to yield photo-active holo-proteins. PYP was incubated with an activated-ester form of *p*-coumaric acid for an hour at room temperature in the dark (204). Cell free extract containing AppA was incubated with an excess of flavin adenine dinucleotide (>95% pure, Sigma-Aldrich) while YtvA was incubated with an excess of riboflavin 5′ monophosphate (>85% pure, Sigma-Aldrich), both for an hour on ice in the dark.

UV-VIS Spectroscopy— Measurement of absorbance spectra and photocycle kinetics were performed on a HP 8453 diode array spectrophotometer (Hewlett Packard). Kinetic measurements were taken with a time resolution of 100 ms after white light photoflash excitation of PYP in 100 mM tris-Cl (pH 8.0). For measurement of the rate of receptor state recovery, AppA and YtvA (in 100 mM tris–Cl, pH 8.0) were irradiated with saturating actinic white light from a Schott KL1500 light source and allowed to revert to the receptor state in the dark. Spectra were recorded every 30 s for 45 minutes (AppA) or every 60 seconds for 120 minutes (YtvA). Absorption changes at 450 nm (YtvA) and 495 nm (AppA) were analyzed by mono-exponential fits to the data using the solver function of Excel (Microsoft corporation, Redmond, USA) in least squares analysis as described before (205). Absorption changes at 446 nm for PYP were analyzed by both mono- and bi-exponential fits to the data to account for the contribution of a slower recovering variant of the protein. However because bi-exponential fits did not yield different or better fitting results, mono-exponential fits to the data are also presented for PYP.

pH-titrations of PYP and azPYP— pH titrations were carried out according to Hoff *et al.* (192) in 10 mM tris-Cl, 100 mM KCl buffer. pK_a values, and *n*-values (or: Hill coefficients) expressing the degree of cooperativity, were calculated by fitting the data to a modified Henderson-Hasselbalch equation [1], in which *n* describes the steepness of the transition by using the solver function of Excel in least squares analysis as described above.

$$pG = \frac{1}{1+10^{n(pH-pK)}} \qquad [1]$$

Sample preparation for Mass spectrometry— For holo-mass measurements proteins were desalted on a C_4 (YtvA) or C_{18} (PYP, AppA) Ziptip (Millipore, Bedford, USA), and eluted in 60% acetonitrile, 0.1% formic acid, 49.9% water prior to mass analysis. For peptide mass fingerprinting and tandem-MS experiments, proteins were digested overnight with 1:50 (w/w) protease:protein, with trypsin gold mass spectrometry grade (Promega, Madison, USA) at 37 °C. Prior to analysis, peptides were desalted with Ziptip C_{18} (Millipore, Bedfort, USA) and eluted in 60% acetonitrile, 0.1% TFA, 49.9% water and diluted 10 times before being loaded on the reversed phase chromatography system. *B. subtilis* lysates were obtained by sonication as described above and were digested overnight with trypsin gold mass spectrometry grade, before azhal peptides were enriched by diagonal chromatography as described in *Chapter 3*. Subsequently, enriched pools were loaded on the LC tandem-MS system described below.

Mass spectrometry— Reflectron MALDI-TOF mass spectra were recorded on a Micromass TofSpec 2EC (Micromass, Whyttenshawe, UK). Holo-masses of proteins were determined by offline spray using Econo12 coated-glass emitters (New objective, Woburn, USA) and a quadrupole time-of-flight mass spectrometer (Q-TOF; Micromass, Waters, Manchester, UK). Samples for LC tandem-MS studies were separated on a reversed phase capillary column (150 mm x 75-µm PepMap C_{18}; LC Packings, Amsterdam, The Netherlands). Sample introduction

and mobile phase delivery at 300 nl/min. were performed using a Ultimate nano-LC-system (Dionex, Sunnyvale, CA) equipped with a 10-µl injection loop. All solvents used are of LC-MS grade (Biosolve, Valkenswaard, the Netherlands). Mobile phase A was water + 0.1% formic acid, and mobile phase B was 50% acetonitrile, 50% water + 0.1% formic acid. For separation of peptides, a two step gradient of 5 – 100% solvent B over 22 minutes was used. Eluting peptides were electrosprayed into a Q-TOF (Waters, Manchester, UK). The most abundant ions from the survey spectrum, ranging from m/z 350 to 1500, were automatically selected for collision-induced fragmentation using Masslynx (Waters, Manchester, UK). Fragmentation was conducted with argon as collision gas at a pressure of 4 x 10^{-5} bars measured on the quadrupole pressure gauge.

Data analysis— Mass spectra of intact proteins were smoothed (Masslynx using a Savitsky-Golay smoothing filter with smoothing window and number of iterations set to 2) and deconvoluted (using the Maxent algorithm) to obtain average masses for the proteins measured. Deconvoluted peak lists from tandem-mass spectrometry experiments were generated by proteinlynx. Next, these pkl-files were combined using notepad. Combined pkl files were submitted to the the MASCOT search engine version 2.1 (Matrix Science, London, United Kingdom) using search parameters as described in detail in *Chapter 3* to identify azhal-containing peptides. A search with these parameters was conducted in a local database of the *B. subtilis* proteome (4238 sequences, 1269237 residues, release 14.2, October 1, 2008, Uniprot consortium). A significance threshold of 0.01 was set, resulting in a threshold score of 38. 'Mudpit scoring' and an ion score cut-off of 38 was applied in order to assure that all assigned peptides had a p-value of <0.01.

3

Identification of newly synthesized *E. coli* proteins by enrichment of azidohomoalanine-labelled peptides using diagonal chromatography

Summary

Measuring protein synthesis and degradation-rates on a proteome-wide scale is a daunting challenge. We present a method to identify several hundreds of newly synthesized proteins in *E. coli* upon pulse-labelling cells with the methionine analogue azhal. Instead of 'click' chemistry to selectively modify the azide moiety of azhal with an affinity handle, this approach enriches labelled peptides by use of selective reduction of the azide moiety to an amine and the subsequent change in retention-time in reversed phase chromatography. We show that in addition to the previously known reduction of the azide-moiety, three other reaction products are induced by tris(2-carboxy-ethyl)-phosphine, 2-mercapto-ethanol or dithiothreitol reacting with azhal-containing peptides or proteins.

Subsequently we applied this reaction to enrich newly synthesized proteins formed during a pulse of 15 minutes. Following digestion of total protein, azhal-labelled peptides are isolated by diagonal reversed phase chromatography. Labelled peptides are isolated by a retention-time shift between the first and second dimension of chromatography that is induced by the selective reaction for the azido group in labelled peptides using tris(2-carboxy-ethyl)-phosphine. Selectively modified peptides enriched by the retention-time shift are identified by tandem mass spectrometry. Accordingly we identified 527 proteins representative of all major Gene Ontology categories in *E. coli* after a pulse of 15 minutes with azhal. The work presented here opens avenues towards the relative quantitation of proteins on a proteomic scale, synthesized during a brief period under different physiological conditions. Such data, in combination with microarray experiments, will enable assessment of the separate contributions of transcription and translation to the regulation of gene expression.

Introduction

Knowledge about protein synthesis and degradation-rates on a proteome-wide scale is an important requirement for advanced modelling of the kinetics of cellular response networks. Pulse-chase labelling with radiolabeled compounds, combined with separation of proteins by two dimensional gel electrophoresis has already been applied (123, 126). However, this approach has drawbacks, such as difficulties to detect very acidic, basic or hydrophobic proteins (e.g. membrane proteins). The possible occurrence of more than one protein in a gel spot, masking the relative contribution of each species to the total radioactivity, is another intrinsic difficulty.

The use of amino acids labelled with stable-isotopes rather than radio-isotopes, is a solution which is applicable to a mass spectrometry based proteome-wide approach (113, 115, 118, 119, 181). However, this method needs extensive labelling times, as the unlabeled bulk of the protein content of the cell will also be detected. Detection of small amounts of labelled, newly-synthesized proteins, in the presence of large amounts of unlabeled proteins is severely limited by the dynamic range of the mass spectrometer. This requirement for longer labelling times hampers identification and quantitation of transient changes in protein expression, following perturbations upon pulse-labelling. What is needed is an amino acid analogue that can be distinguished from its natural counterpart, can be used in a gel-free proteomics approach, and will facilitate the isolation of newly synthesized proteins from a large pool of pre-existing ones. This will enhance identification and increase the dynamic range as well as the sensitivity of detection for transiently expressed proteins.

In recent efforts, non-natural amino acids have been used to distinguish between newly synthesized proteins and pre-existing ones. Azhal, a methionine analogue, was reported to be efficiently incorporated into recombinant proteins as well as protein complexes produced in methionine-auxotrophic *E. coli* strains and in mammalian cells grown in its presence (143-151). Azido groups can be selectively modified, by application of so called 'click' chemistry using the azide's reactivity towards chemical moieties such as alkynes with copper catalyzed (3+2) cyclo-addition (148, 152), or octynes with the strain-promoted cyclo-addition (154-156, 165) and phosphines through the Staudinger ligation (144, 159-161). Use of these reactions enables fluorescent labelling of newly synthesized proteins, which allows for example following the movement of newly synthesized proteins through different cellular compartments (162, 164, 165). In addition, these specific and efficient chemical reactions also open up the possibility to selectively purify newly synthesized proteins or peptides that are labelled with azhal by attachment of an affinity handle (167, 168, 170). Therefore, azhal seems a promising label with which to probe protein synthesis- and degradation-rates on a proteome-wide scale.

Notwithstanding the promise of the affinity purification approach of azido-proteins and peptides, we present an alternative for enrichment of azido-labelled peptides from unlabeled ones. Our approach is based on combined fractional diagonal chromatography (COFRADIC), originally used by Gevaert *et al.* in a proteome-wide approach to identify

methionine containing peptides (213) and also applied to enrich other peptide subpopulations (214-219). Gevaert *et al.* employed oxidation of methionine residues to induce retention-time shifts in reversed phase high performance liquid chromatography (RP-HPLC), in order to sequester methionine containing peptides from non-methionine containing ones. We adapted this approach to isolate azhal-labelled peptides by a retention-time shift. We make use of the azide's susceptibility to reduction by phosphines (158, 161) and induce a retention-time shift by reaction of tris(2-carboxy-ethyl)-phosphine (TCEP) with azhal. This results in the conversion of the azhal residue into a di-aminobutyrate residue (**3** in Figure 1), via the reduction of the azido-group to an amino-group. The more polar nature of an amino group compared to an azido-group induces shifts on RP-HPLC (Figure 1a). However, in addition to the already described reduction of azide's by phosphines and (di)thiols (220,

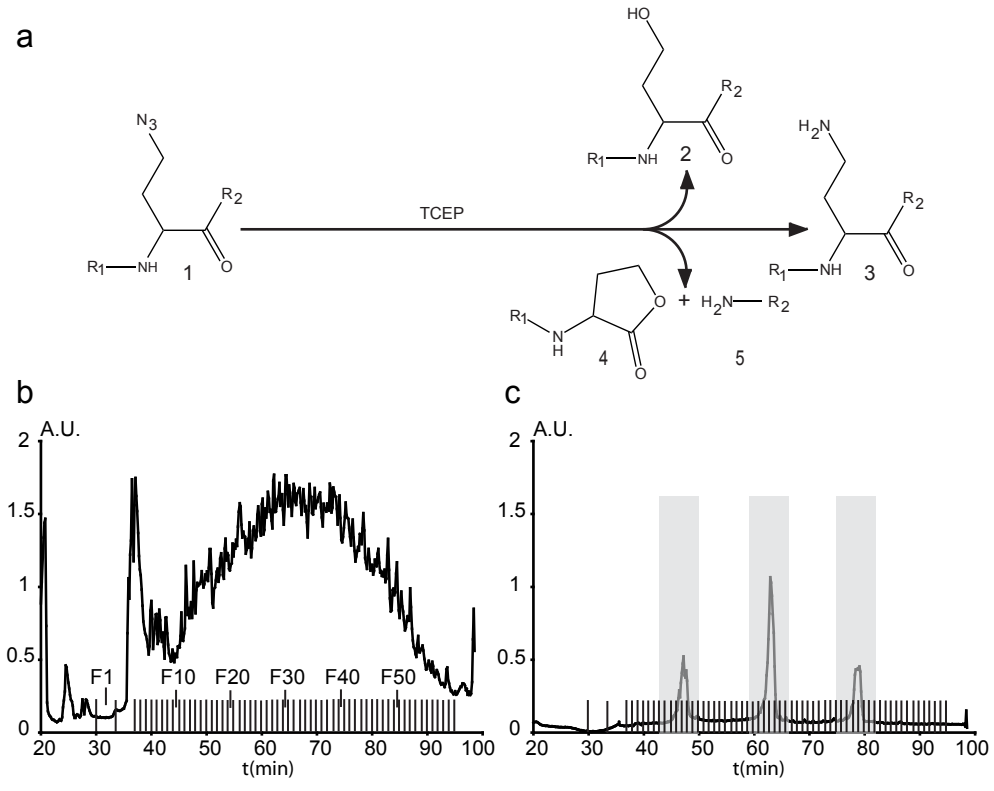


Figure 1. **The COFRADIC approach**. reactions products resulting from an azhal (**1**) -containing peptide reacting with TCEP are: a peptide containing either a homoserine residue (**2**), or a di-aminobutyrate residue (**3**), or an N-terminal cleavage product with a homoserine lactone residue at its C-terminus (**4**) and a C-terminal product with a normal N-terminus (**5**) due to specific cleavage at the azhal residue (**a**). Primary RP-HPLC run of a tryptic digest of a labelled *E. coli* proteome; a total of 65 one-minute fractions are collected from 30 to 95 minutes. In order to diminish the number of secondary runs to 16, primary fractions (three at a time, see Table IV) separated by an interval of 16 minutes are pooled (**b**). Pooled fractions are treated with TCEP and run again under identical conditions. Unlabeled peptides are not modified and will run at the same retention-time; while labelled peptides react with TCEP and shift their retention-time. The non-shifted fractions plus 3 adjacent fractions on the front and back are discarded (grey area in **c**). The remaining fractions containing shifted peptides are pooled, lyophilized and analyzed by tandem-MS.

221) we found through our investigations into Cu^{I}-catalyzed (3+2) cyclo-additions of several different alkynes to azhal-containing peptides (with the in situ reduction of Cu^{II} to Cu^{I}) that two competing reactions are induced. Here we show that these competing reactions lead to three additional reaction products (**2**, **4** and **5** in Figure 1) and demonstrate the ability of this method, i.e. azido-peptide isolation by COFRADIC, to enrich azhal-labelled peptides and thus identify newly synthesized proteins in a proteome-wide approach. We present data on 527 *E. coli* proteins synthesized during a 15 minute pulse.

Results

Two additional reactions are induced by TCEP in azhal containing peptides—In our investigations into Cu^{I}-catalyzed (3+2) cyclo-additions of several different alkynes to peptides containing azhal, with the in situ reduction of Cu^{II} to Cu^{I}, we discovered smaller fragments aside from the expected products. Control experiments in which Cu^{II} or the alkyne moiety, or both, were omitted showed a marked increase in peptidolysis in the presence of reducing agents. And so these reducing agents were chosen to investigate the underlying mechanism of cleavage. Two products were formed when the model octadecapeptide **Pan016** (sequence: PPHHHHHHPPRGFGAzGFR) was incubated in buffers containing either TCEP, 2-mercapto-ethanol (2ME), or dithiothreitol (DTT). As expected, a peak corresponding to the reduction of azhal to 2,4- di-aminobutyrate at m/z 2108.0 was observed under these conditions. Surprisingly, an additional product at m/z 1729.8 was observed. Low-energy collision-induced dissociation (CID) of this product in an ESI-Q-Fourier Transform mass-spectrometer (FTMS) revealed an N-terminal fragment of Pan016 that resulted from cleavage of the peptide bond C-terminal to azhal together with the loss of the azide group.

To test the general occurrence of this cleavage reaction in a protein that contains azhal instead of methionine, we used recombinant PYP from *H. halophila* produced in a methionine-auxotrophic *E. coli* grown on media containing azhal. A pure preparation azPYP, in which all six methionine residues have been replaced by azhal, was obtained (see *Chapter 2*). Reduction and cleavage of purified azPYP by TCEP, 2ME, or DTT was analyzed by gel electrophoresis and mass spectrometry (Figure 2 a, b). It was found that TCEP was able to cleave the protein under all conditions investigated, whereas both the rate and yield of the cleavage with DTT and 2ME increased drastically under denaturing conditions (4M urea). Fragments N-or C-terminal to all positions of azhal in reductively treated azPYP were detected (Figure 2c). All fragments, except for the C-terminal peptide ending in valine, have a C-terminal residue with a mass of 83 Da which corresponds to azhal minus a HN_3 moiety, as confirmed by CID. Peptides with 'missed cleavages' have masses in accordance with a reduction of their internal azhal-residues. The fact that reduction-induced cleavage occurs C-terminal to all positions of azhal in azPYP indicates that the cleavage reaction apparently poses no special demands on the residue C-terminal to the scissile bond.

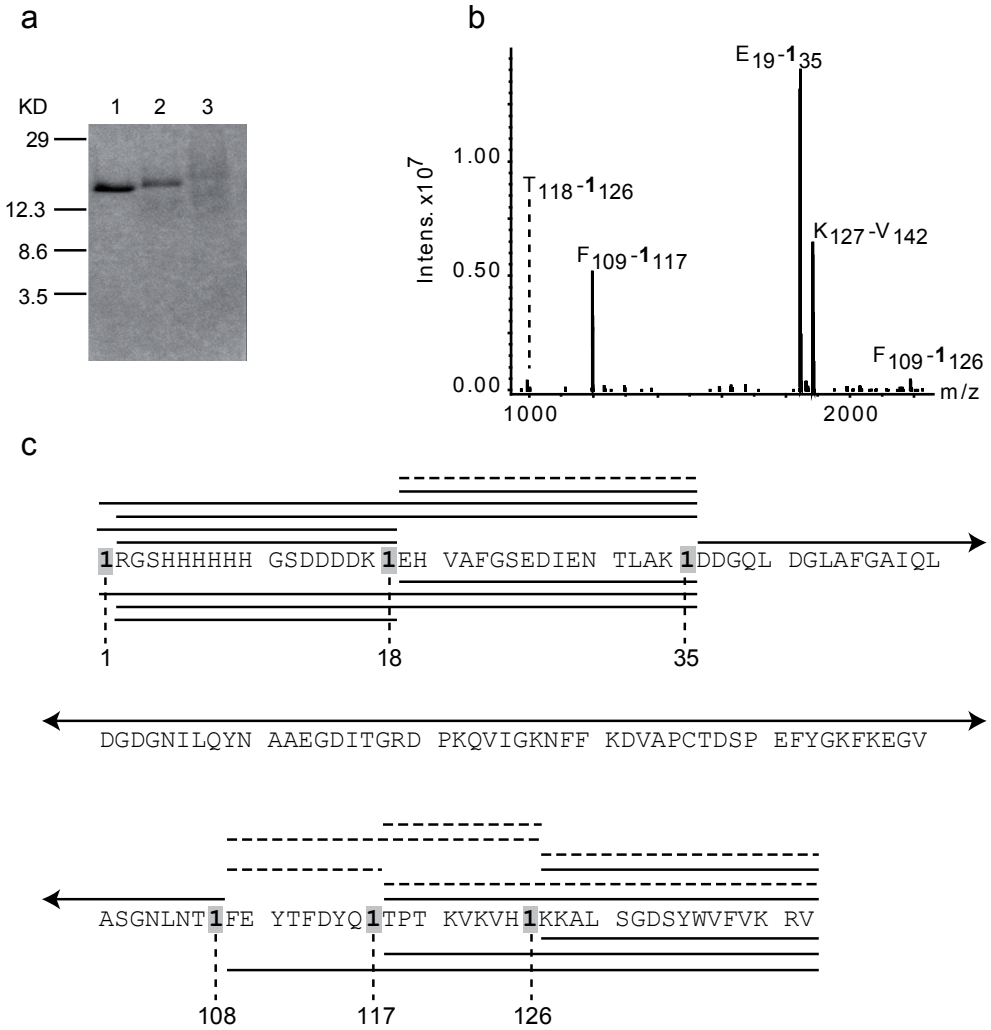


FIGURE 2. **Recombinant azPYP is cleaved by TCEP.** Coomassie Blue stained gel containing equal amounts (23 mg) of azPYP incubated without (lane 1) and with 10 mm (lane 2) and 100 mm (lane 3) TCEP at pH 5 (**a**); the intact protein disappears and is cut into smaller fragments that do not nicely resolve on the polyacrylamide gel. Deconvoluted ESI FTMS spectrum of TCEP-cleaved azPYP (**b**). Sequence of His-tagged PYP from *H. halophila* (**c**). Intact molecular mass and tryptic peptide mapping confirmed >95% incorporation of azhal (**1**) at methionine coded residues. A coverage map obtained with TCEP (above the sequence) or DTT (below the sequence) is presented. Solid lines indicate peptides that result from cleavage that were detected with MALDI-TOF MS; dashed lines indicate peptides that were detected with ESI-FTMS.

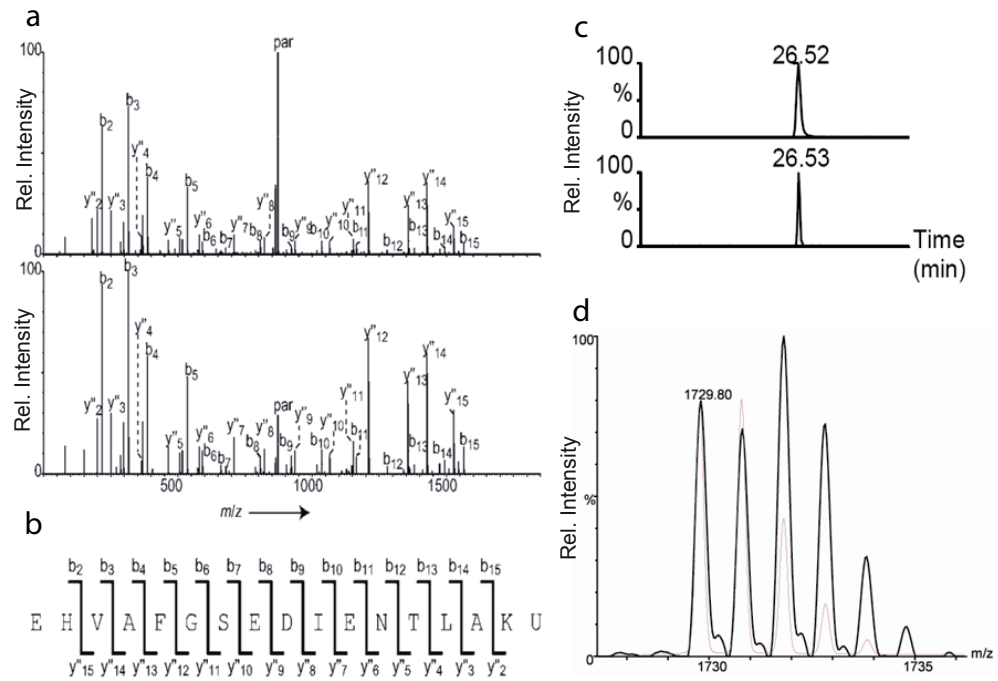


FIGURE 3. **Cleavage by TCEP results in a C-terminal homoserine lactone moiety.** Comparison of the cleavage of azPYP with TCEP and methionine containing PYP with cyanogen bromide, charge-deconvoluted tandem-MS fragmentation spectra of the doubly charged ion of m/z 921.9 that eluted at identical retention-times with annotation of fragment ions and parent ion (par) azPYP × TCEP (top spectrum) and PYP × CNBr (lower spectrum); identical fragment spectra verify the equivalence of both peptides (**a**). Sequence of the peptide and annotation of the retrieved fragment ions. U=homoserine lactone (**b**). Digests loaded onto an LC tandem-MS system (see *experimental procedures*), extracted ion chromatograms of the doubly charged signal at *m/z* 921.9 of azPYP × TCEP (top trace) and PYP × CNBr (lower trace) show less than 1 sec retention-time deviation (**c**). Reductive cleavage by TCEP of peptide Pan016 performed in 50% ^{18}O enriched water, yielding an isotopic pattern that is indicative of the incorporation of one oxygen atom from water (black trace). The simulated isotope pattern of the peptide with elemental composition $C_{79}H_{104}N_{30}O_{16}$, assuming natural abundance isotopes (grey trace). A 1:1 ratio of ^{18}O incorporation in the cleavage experiment can clearly be inferred (**d**).

Next, we carried out cleavage of peptide Pan016 by TCEP in 50% ^{18}O-enriched water, which confirmed that the peptide bond C-terminal to azhal is cleaved by hydrolysis, and indicated that only one ^{18}O atom is incorporated into the resulting C-terminal residue of the N-terminal peptide fragment (see Figure 3). The C-terminal residue is considered to have a homoserine lactone structure based on the following observations:

i) When the cleavage product of peptide Pan016 was incubated under basic conditions, the mass of the product increased by 18 Da, which was added to the C-terminal residue as confirmed by CID; this addition could be reversed by incubation for one hour in anhydrous trifluoroacetic acid (TFA) (222).

ii) Incubation of the cleaved product of peptide Pan016, or the cleaved products of azPYP in butylamine, led to full conversion with an irreversible increase in the mass of the cleavage products by 73 Da, which is consistent with the addition of a C-terminal butylamine group (222), as verified by CID.

iii) It was found that both RP-HPLC retention-times and CID tandem-MS analysis (see Figure 3) of the presumed carboxyterminal homoserine lactone peptides from TCEP-cleaved azPYP and the corresponding carboxyterminal homoserine lactone peptides from cyanogen bromide (CNBr) cleaved PYP were identical (see Figure 3).

With TCEP, the reaction runs to completion in 100 minutes at 50 °C and pH 5, based on the disappearance of starting material. The ratio of cleaved/reduced product varied little as a function of pH value (a range of pH 3–11 was tested) or temperature, as quantified from RP-HPLC followed by MALDI-TOF MS identification of the collected peak fractions. A maximized ratio at room temperature was accomplished at pH 5, resulting in 55% (±5%) cleavage, whereas at pH 9 approximately 40% (±5%) was cleaved (sum total of homoserine and homoserine lactone cleavage products).

In the course of this study we became aware of the formation of a fourth TCEP-induced reaction product of azhal-containing peptides next to the two cleavage products and di-aminobutyrate described above. This reaction leads to a –25 Da shift of the peptide nominal molecular mass and a slightly smaller retention-time shift than di-aminobutyrate (Figure 4). Inspection of tandem mass spectra of Δm –25 Da peptide ions shows that the structural modification is localized on the azhal residue (Figure 4). The residue mass of this reaction product is 101.05 Da. This mass is odd thus the reaction product contains one or three nitrogen atoms. The mass difference of -25 Da excludes the latter and points to replacement of the azide functionality by a hydroxyl group, giving rise to a homoserine moiety (Figure 1). The proposed reaction scheme of this conversion and those of cleavage and reduction are presented in Figure 5.

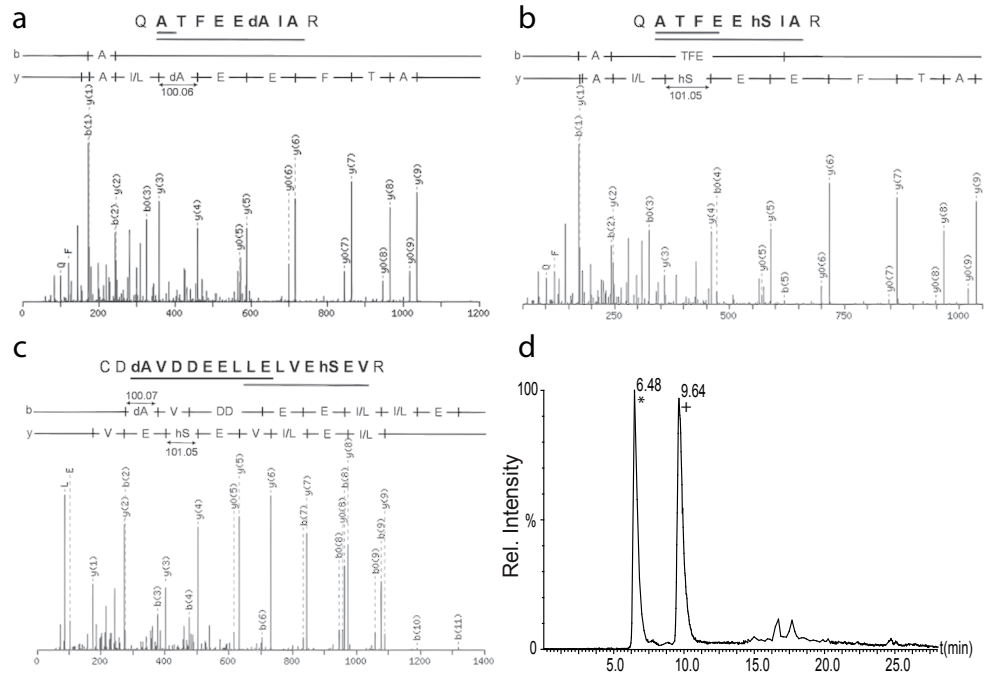


FIGURE 4. **The homoserine product is localized at the azhal-residue and has a different shift in retention-time compared to di-aminobutyrate.** Tandem-MS spectra of a peptide observed in two forms**.** The azhal residue can be reduced, forming di-aminobutyrate (dA), or react to form homoserine (hS) as depicted in (**a**) and (**b**) respectively. Example of a peptide containing two azhal residues with one modified to di-aminobutyrate and the other to homoserine (**c**). Spectra pinpoint the localization of the homoserine modifications on the azhal residue of the peptide. Extracted ion chromatogram of a peptide identified in two forms, illustrating the retention-time difference on nano RP-HPLC between a peptide with a di-aminobutyrate residue (*) or a homoserine residue (+) as detected by ESI Q-TOF (**d**). The average retention-time difference between peptides containing either residue is 3 minutes with a standard deviation of 35 seconds, the di-aminobutyrate form eluting earlier than the homoserine form.

Isolation of azhal-containing peptides by COFRADIC— Although *E. coli* does not sustain long term growth on azhal-containing media, incorporation is normal for the first 30 minutes. Furthermore, with no obvious toxic side effects, this amino acid analogue seems well suited to label and identify newly synthesized proteins over short periods of time (see *Chapter 2*). To test this, azhal was used to identify proteins synthesized by *E. coli* during a 15 minute pulse period in two separate cultures to take biological variation into account. The percentage of labelled protein is estimated to be 8~9% of total protein after 15 minutes of growth based on increase of OD_{600}. After digestion, labelled peptides were isolated using the COFRADIC approach as described in the experimental procedures and depicted in Figure 1.

The experimental setup is optimized to sequester di-aminobutyrate-containing peptides. Peptides in which the azido-group has been converted by TCEP into an amine-group generally elute 3-7 minutes earlier than their parent compound (223). Therefore, they are expected to be collected in the off-diagonal fractions, as are the homoserine-containing peptides which have a smaller retention-time shift compared to the

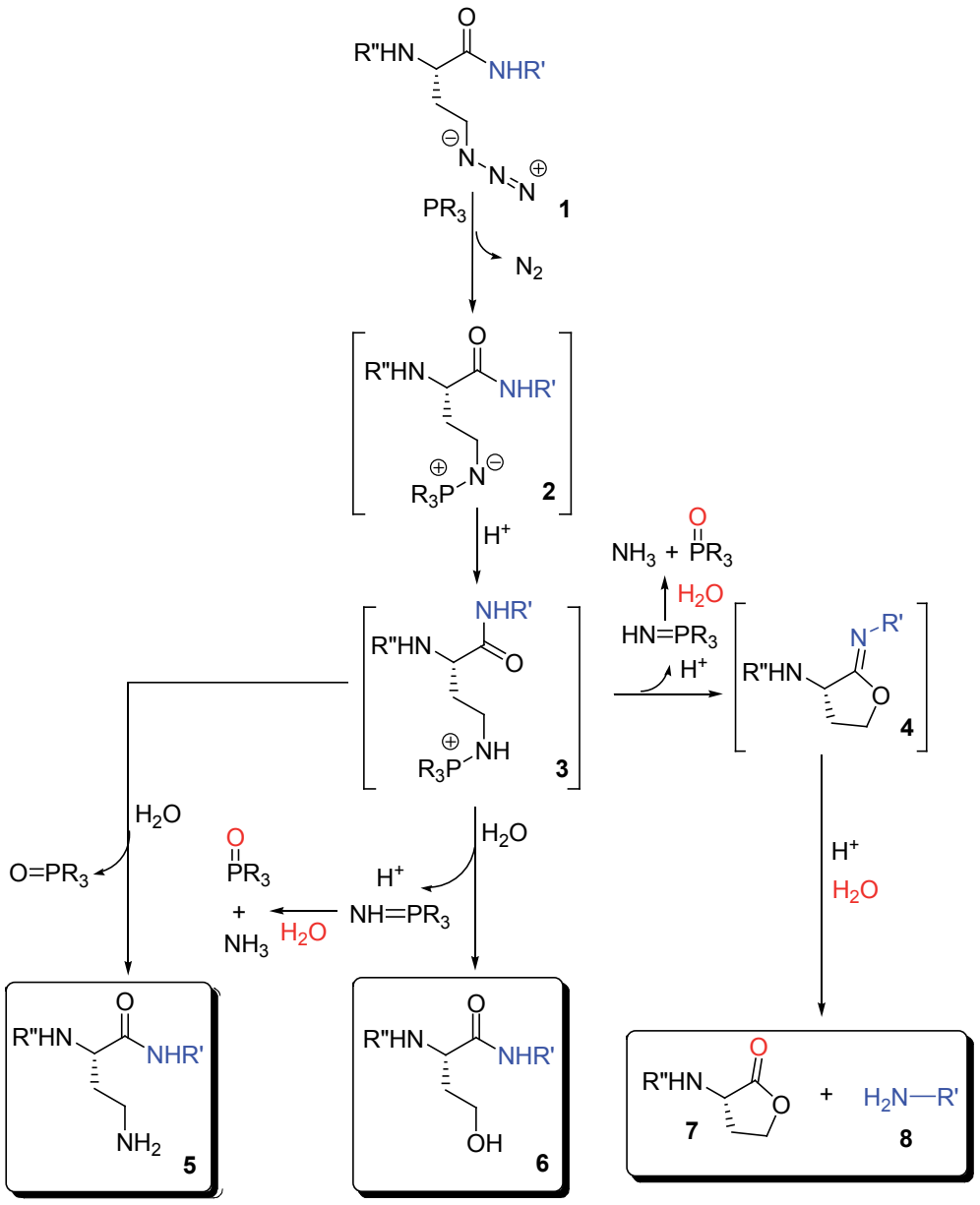


FIGURE 5. **Proposed reaction scheme of reduction and cleavage of azhal containing peptides by TCEP.** Scheme modified from Back *et al.* (143), PR_3=TCEP, which induces competing reactions on azhal-labelled peptides (1). The homoserine product (6) is formed by nucleophilic displacement of the iminophosphorane (3) by H_2O. Also shown is the product of reduction, di-aminobutyrate (5), and the cleavage products (7 and 8). Note that there is an alternate route to cleavage products 7 and 8 as well as to reduction product 5 as reported by Back *et al.* (143).

di-aminobutyrate-containing peptides. In contrast, TCEP-induced cleavage products have a much broader range of retention-time differences as compared to their parent compounds, and will elute both in on- and off-diagonal fractions. Because of their relative small size, cleavage products are considered less useful for protein identification purposes than di-aminobutyrate containing peptides. Correct assignment of small peptides by MASCOT is difficult. Moreover, small peptides are more often singly charged, thereby escaping selection for tandem-MS. To ensure that the homoserine and di-aminobutyrate are the predominant products formed, we chose a 50 mM pH 8 buffer to perform the overnight reaction in. The distribution of the reaction products under these conditions is given in Table I for the model peptide Pan016 and shows that indeed the reduction and formation of the hydroxyl predominate.

TABLE I
Relative distribution of reaction products of azhal-containing peptide Pan016, induced by TCEP

TCEP-induced reaction product	% of total	s.d.
DAB-reaction product	42	2
HS-reaction product	37	2
N- and C- terminal cleavage products	21	3

A total of 2709 peptides were identified in the fractions containing reaction products from the replicate cultures. Of these peptides, 1663 were identified as reaction products derived from azhal-containing peptides, i.e. peptides from newly synthesized proteins. Because of the setup of the COFRADIC approach, the number of identified cleavage products is less than the number of identified peptides in which the azhal residue has been converted to a di-aminobutyrate or homoserine residue (Table II). Also unlabeled peptides were identified; the presence of unlabeled species in the off-diagonal fractions is probably due to peak broadening of on-diagonal material (Figure 1c). The contribution of non azhal-containing peptides in the off-diagonal pools is minimized by discarding both the on-diagonal fraction and three adjacent fractions on the front and back. Discarding even more fractions would lead to loss of shifted reaction products.

Newly synthesized proteins identified by azhal labelling— From the two biological replicates 527 proteins were identified exclusively using reaction products from azhal-containing peptides as described in the experimental section. Of these, 294 were found in both replicate cultures. The false positive rate for all proteins identified in the two experiments is under two percent. For proteins identified in both experiments this rate even drops to less than half a percent. To investigate how the proteins synthesized during the 15 minute pulse period are distributed according to function and location, they were mapped to their Gene Ontology terms using GO-Miner. Of all 527 newly synthesized proteins identified in the two samples,

TABLE II
Number of identified TCEP-induced reaction products

TCEP-induced reaction product	number of peptides
DAB-containing peptides	641
HS-containing peptides	644
HS and DAB-containing peptides	73
N-terminal fragment of cleavage at azhal	22
C-terminal fragment of cleavage at azhal	198
DAB- or HS-containing peptides combined with cleavage at Azhal*	85

TCEP-induced reaction products in a tryptic digest of an *E. coli* proteome pulse-labelled with azhal were isolated by COFRADIC and identified by LC tandem-MS as described in Experimental procedures. *DAB- or HS-containing peptides and N- or C-terminal cleavage products derived from the same multiple azhal-residue containing precursor peptide.

497 can be mapped to GO terms. The proteins newly synthesized during the pulse period are distributed over all the major categories of Gene Ontology terms present in the *E. coli* proteome (Table III). All major pathways are represented, including energy metabolism, transcription, translation, cell cycle, signal transduction, stress response and taxis. This demonstrates again that normal cellular translation of proteins involved in the major cellular pathways continues in the presence of azhal, and that azhal is incorporated into these proteins during pulse-labelling. That means that these processes are relatively unaffected, at least at the level of translation.

Upon closer inspection of the proteins related to energy metabolism, we find 14 proteins involved in glycolysis and 11 proteins which are part of the Krebs cycle. Furthermore three proteins are part of the electron transport chain and five are involved in proton translocation coupled to ATP synthesis. Also proteins involved in transcription of genes and regulation thereof are well represented, with 39 proteins involved in transcription and 24 of those having transcription regulatory functions. Furthermore the translational machinery is labelled during the pulse period with 36 out of the 56 ribosomal proteins detected next to 21 proteins involved in tRNA amino acylation.

Cells were labelled in minimal medium containing 19 natural amino acids. However, proteins involved in biosynthesis of lysine, serine, arginine, glutamine, cysteine and methionine are still being made, as witnessed by azhal incorporation. The strain used is an auxotroph which has deletions in the genes encoding MetE, MetH and YagD which catalyze final steps in the synthesis from different precursors to methionine. Of these three, MetE is detected, however since it has a 391 amino acid deletion it is not able to synthesize methionine (180). Three MetE peptides were identified, two of which are found in the N-terminal part from residue 1 to 164 before the deletion and one in the C-terminal part from residue 165 to 362 after the deletion. The observation that proteins involved in methionine biosynthesis are

TABLE III
Newly synthesized proteins assigned to Gene Ontology

Description	Total number of proteins in category	Number of newly synthesized proteins	Relative ratio*
cell cycle	58	11	1.2
signal transduction	150	16	0.7
response to stress	91	12	0.8
taxis	22	4	1.1
transport	765	66	0.5
electron transport	247	27	0.7
transcription	359	39	0.7
translation	110	74	4.1
DNA metabolic process	187	17	0.6
RNA metabolic process	457	72	1.0
protein metabolic process	313	115	2.3
protein catabolic process	3	1	2.0
protein modification process	55	2	0.2
lipid metabolic process	153	26	1.0
secondary metabolic process	31	10	2.0
generation of precursor metabolites and energy	287	44	0.9
ribosome biogenesis and assembly	22	5	1.4
membrane organization and biogenesis	5	2	2.5
cell wall organization and biogenesis	47	5	0.7
intracellular	504	162	1.7
cytoplasm cytosol	300	138	2.5
membrane	1144	80	0.4
periplasmic space	168	27	0.9
cell wall	115	2	0.1
outer membrane	90	19	1.1

Identified proteins are assigned to Gene Ontology categories of biological processes and cellular localization by GO-miner as described in the experimental section.*Relative ratio of the GO annotations per category in the protein dataset over those in the proteome. Ratio>1 indicates a relative over-representation of the category in the dataset compared to the proteome, ratio<1 relative under-representation.

being made does not allow the inference that this reflects the cells reaction to the methionine deficiency during labelling with azhal, as many other amino acid biosynthesis routes (see above) are also turned on.

Additional extraction methods are needed to isolate membrane proteins besides urea extraction. In our approach, no additional extraction steps were used for membrane proteins. Thus an under-representation of membrane proteins is expected in our data set. Table III shows this indeed to be the case. The relative number of membrane proteins identified in comparison to the total set of proteins identified is similar to the number found by Gevaert *et al.* with a similar protein extraction protocol as used here (Figure 6). In another study more membrane

proteins were identified, because of a separate extraction of membrane proteins from cellular debris obtained after extracting soluble protein (224).

A complication with regard to the replacement of methionine by azhal is related to methionine's role in methylation processes. The formation of s-adenosyl-methionine (SAM), a donor of methyl-groups in the cell, from methionine, is catalyzed by the metK gene product. As is immediately obvious from its structure, azhal cannot replace methionine as methyl donor in the cell (see *Chapter 1*). There could thus be a concern that methylation reactions go awry during the short labelling with azhal because of lack of SAM. Indeed METK is identified among the newly synthesized proteins which could indicate that cells sense a lack of SAM and initiate its synthesis. However METK is also amongst the identified proteins in two other studies of *the E. coli* proteome not using azhal (213, 224). This means that the presence of METK not necessarily points to an azhal-induced lack of SAM.

An additional effect of a lack of SAM during pulse-labelling with azhal could be a lack of DAM mediated DNA methylation and subsequent induction of the SOS response (225). However, only one out of twenty SOS response gene products was identified. Although this could point to an early response or a low level of steady state synthesis normally present in the cell, other studies of the *E .coli* proteome identified multiple SOS response gene products as well (213, 224). Therefore, it cannot be concluded that the SOS response gene products

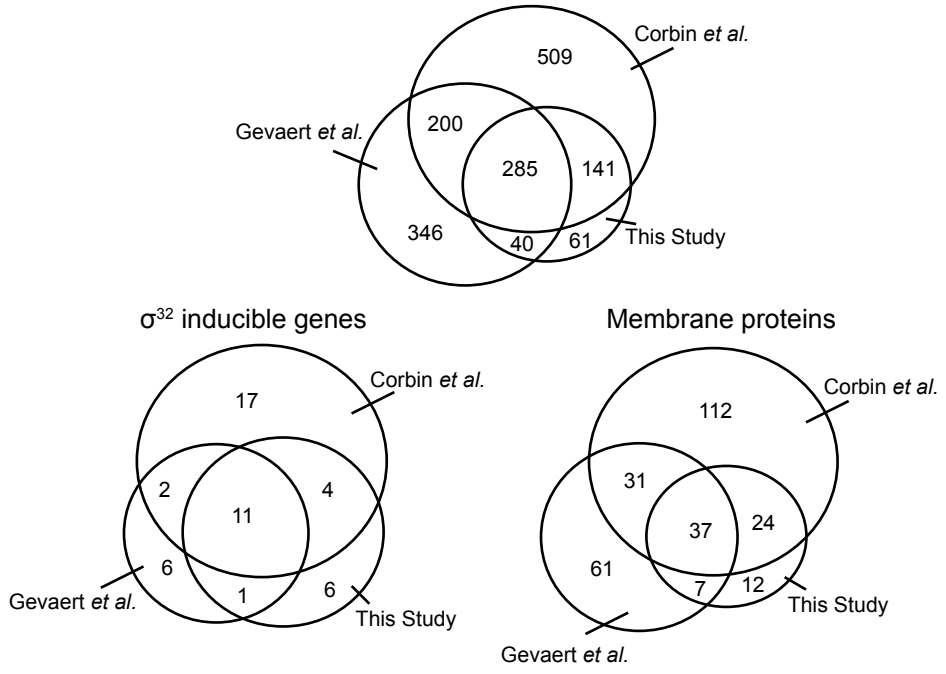


FIGURE 6. **Venn diagrams of proteins identified in this and alternative studies.** Overlap between proteins identified in this study and earlier studies of the *E. coli* proteome using 2D-LC by Corbin *et al.* (224) and COFRADIC, Gevaert *et al.* (213). Overlap between membrane proteins as well as σ^{32} (RP32) inducible gene products are shown.

are induced by a possible lack of methylation of DNA synthesized during pulse-labelling with azhal.

Another concern is that azhal-labelling could influence protein folding and function, and induce a "heat shock" response to unfolded proteins. Heat shock-related chaperone systems are induced by the *rpoH* gene product. This alternate sigma factor is induced by heat shock and growth at higher temperatures during exponential aerobic growth (226-228). Indeed the *rpoH* gene product σ^{32} was identified amongst the newly synthesized proteins and we wondered if azhal might induce a heat shock response due to misfolding of proteins. Large scale studies with respect to σ^{32} inducible genes have been carried out in which 97 genes were identified to be part of the σ^{32}-regulon (229, 230). From this σ^{32}-regulon, 22 gene products were identified amongst which two chaperone systems, i.e. the DnaJ, DnaK, GrpE chaperone system and GroEL. However, comparison of our results with two other studies of the *E. coli* proteome (213, 224) shows that a comparable number of σ^{32} inducible gene products were identified in these studies.

DISCUSSION

A new method to identify newly synthesized proteins by the use of azhal as a pulse-label is presented. It is based on the induction of a selective retention-time shift in azhal-containing peptides between two reversed phase chromatographic runs to enrich these peptides. The reaction is the selective reduction of the azide-moiety of azhal to an amine which can be induced by phosphines such as TCEP but also by (di)thiols such as 2ME and DTT. We found that apart from reduction to an amine two competing pathways are also induced. In one pathway the azhal residue is modified to a homoserine residue (Figure 1, **2**) where the azide function is substituted by a hydroxyl group resulting in a smaller induced retention-time shift compared to the di-aminobutyrate residue. In the other pathway the peptide bond at the C-terminal side of an azhal residue is cleaved. In the N-terminal product of this cleavage reaction the C-terminal azhal has been converted into homoserine lactone (**4**); the C-terminal cleavage product is an unmodified peptide (**5**). We used this approach to identify reaction products stemming from azhal-containing peptides from a digest of proteins isolated from *E. coli* that were pulse-labelled with azhal for 15 minutes and found that we were able to identify 527 proteins newly synthesized during pulse-labelling.

Growth rate and viability in the presence of azhal or methionine is the same during the first 30 minutes, however, after prolonged labelling growth arrest does occur (*Chapter 2*). This can be explained by assuming dysfunctioning of essential proteins with crucial roles for one or more methionine residues or as a result of azhal induced misfolding of such proteins. An important prerequisite for a pulse-label is that it does not cause major changes in protein expression during the labelling period. The labelled proteins should be representative of the translational activity of the cell at the start of the pulse. No major differences were found between our dataset of newly synthesized proteins and the lists of proteins identified by others (213, 224) in *E. coli* grown under comparable conditions in the absence of azhal. Moreover,

there is evidence for normal protein processing of the N-terminal azhal residue in *E .coli* (175). In addition, evidence for mostly normal localization and folding of azhal-labelled proteins comes from the observation that the membrane protein OMPC is displayed at the cell surface of *E. coli* (148-150), and recombinantly produced virus-like particles assemble normally (151) when methionine is replaced by azhal. Also the results presented in *Chapter 2* for the functionality of three photo-active proteins in which methionine is replaced show that function is largely maintained. These findings, in combination with the fact that over 500 proteins are found to be labelled representing all major pathways in the cell, lead us to believe that azhal is a suitable label for pulse-chase experiments in *E. coli*.

Our COFRADIC-based method to isolate labelled peptides is an alternative to using affinity purification either by a protein centric approach (167, 168) or a peptide centric approach as described by us more recently (170). It has the advantage that it utilizes standard chromatographic techniques and chemicals to achieve sequestration of labelled from unlabeled peptides. Furthermore, it provides a simple and robust approach to enrichment of labelled peptides. The TCEP-based chemistry is compatible with a range of pH, temperature and buffer compositions. Apart from purification this method also provides chromatographic fractionation of labelled peptides, thereby facilitating mass spectrometric identification. We clearly demonstrate the power of azido-peptide isolation by COFRADIC to sequester labelled peptides by the large number of proteins identified. Furthermore, we show that only short labelling times are required, half of that used for stable-isotope approaches to pulse-labelling described in *E. coli* (115). We expect that even shorter labelling times are possible. However, contamination with unmodified peptides from main chromatographic peaks, as described above, will limit reduction of labelling times, at least when aiming for identification and quantification of hundreds of proteins.

Altogether, the fact that no severe azhal-related disturbances were observed, combined with the large amount of newly synthesized proteins identified, makes azido-peptide isolation by COFRADIC an excellent tool for the identification of transient changes in protein synthesis, following adaptations to changes in the environment of *E. coli.* The further application of extraction protocols geared towards membrane proteins, should give this approach the ability to measure the changes in synthesis of this class of proteins as well. In addition, the use of stable-isotope labels such as iTRAQ will allow proteome-wide determination of changes of relative rates of protein synthesis, e.g. following a perturbation. The method presented can open up new avenues in systems biology research, by filling the gap of information between transcriptomics and proteomics and generating input for advanced modelling of cellular networks.

Experimental Procedures

Synthesis of L-azhal— L-azhal was synthesized from L-Boc-2,4-di-aminobutyric acid (L-Boc-DAB, Chem-Impex, Wood Dale, USA) by diazotransfer (198) using Triflic azide (TfN_3) as previously described (141).

Peptide synthesis and recombinant protein expression— Following by Boc removal with dioxane/HCl in dichloromethane azhal was reprotected by Fmoc. Fmoc-azhal was used to synthesize peptide Pan016 (sequence:PPHHHHHHPPRGFG**Az**GFR) using standard Fmoc chemistry (Service XS, Leiden, the Netherlands). Azhal-labelled His-tagged recombinant photo-active yellow Protein (PYP) from *H. halophila* was produced as described before (see *Chapter 2*). Briefly: *E. coli* strain CAG18491 transformed with pREP4 and pHISP (199), was grown in M9 medium containing 400 mg/l azhal, 50 µg/ml kanamycin/ampicillin. Protein expression was induced by 1 mM IPTG. After lysis, the chromophore p-coumaric acid was inserted (204) and azPYP was purified on Ni-NTA agarose (Qiagen, Venlo, the Netherlands). Methionine containing His-tagged PYP was produced similarly from transformed *E. coli* CAG18491 grown on M9 medium containing 60 mg/l methionine.

Peptide and protein cleavage with TCEP and cyanogen bromide—cleavage of Pan016 and azPYP was achieved in 50 mM Na-acetate buffer pH 5, and up to 100 mM of TCEP, unless otherwise stated. For ^{18}O incorporation, an aliquot of $H_2{}^{18}O$ (>95% atom 18O, Campro Scientific,Veenendaal, The Netherlands) was added so that water with 50% 18O content was obtained. Reactions were left at room temperature overnight. Homoserine lactones were hydrolyzed by the addition of an excess of unbuffered 1M Na_2CO_3 which was left for 24 h. After vacuum centrifugation, reformation of the lactone was performed in anhydrous TFA for 1h. Peptide and protein cleavage with thiols: cleavage of peptides and protein was performed in 100 mM of either 2ME or DTT buffered by sodium carbonate pH 9.2. To unfold azPYP urea up to 4M (final concentration) was added.

Table IV
Pool scheme primary run COFRADIC

Pool	Fractions	Pool	Fractions
A	7,23,39	I	15,31,47
B	8,24,40	J	16,32,48
C	9,25,41	K	17,33,49
D	10,26,42	L	18,34,50
E	11,27,43	M	19,35,51
F	12,28,44	N	20,36,52
G	13,29,45	O	21,37,53
H	14,30,46	P	22,38,54

Fractions as shown in Figure 1

Elongation experiments—One-pot one stage C-terminal elongation experiments were carried out in 100 mM butylamine pH 9. In one pot two stage experiments the peptide was initially incubated in 100 mM NaAc pH 5 with 100 mM TCEP, and after 8 hours butylamine and NaOH were added so that the concentration butylamine was 100 mM and the pH was elevated to 9. For two step derivatization peptides were harvested by strong cation exchange on Vivaspin-S microcentrifuge tubes (Sartorius, Göttingen, Germany) that were eluted with 200 mM Na_2CO_3. The eluate was acidified with TFA, dried in a vacuum centrifuge and the pellet was acidified again with TFA and once more dried. Subsequently, amines were added either neat or dissolved in dry MeOH to the dried lactone peptides. Methionine containing PYP was cleaved by incubation of 2.5 µg of His-tagged PYP in 20 mg/ml cyanogen

bromide in 70% TFA for 2 hours at room temperature. Prior to mass analysis peptides were desalted with Ziptip C_{18} (Millipore, Bedford, USA), according to the manufacturer's protocol.

Chromatography and mass spectrometry— HPLC separation of cleaved products was performed on a Jupiter Proteo C_{12} column (Phenomenex, Torrance, USA); the gradient of water/acetonitrile being delivered by a SMART system (AmershamPharmacia, Uppsala, Sweden). Reflectron MALDI-TOF mass spectra were recorded on a Micromass TOFSpec 2EC (Micromass, Whyttenshawe, UK). ESI-FTMS and MALDI-FTMS spectra were acquired on a 7T APEX-Q FTMS (Bruker Daltonics, Bremen, Germany) equipped with a CombiSource. For low energy CID the ions were activated in the external collision cell or produced by SORI-CID in the FTMS cell. Samples for LC tandem-MS studies were separated on a reversed phase capillary column (150 mm x 75-µm PepMap C_{18}; LC Packings, Amsterdam, The Netherlands). Sample introduction and mobile phase delivery at 300 nl/min. were performed using a Ultimate nano-LC-system (Dionex, Sunnyvale, CA) equipped with a 10-µl injection loop. All solvents used are of LC-MS grade (Biosolve, Valkenswaard, the Netherlands). Mobile phase A was water + 0.1% formic acid, and mobile phase B was 50% acetonitrile 50% water + 0.1% formic acid. For the separation of peptides, a two step gradient of 5 – 100% solvent B over 22 minutes was employed. Eluting peptides were directly electrosprayed into a Micromass Q-TOF mass spectrometer (Waters, Manchester, UK). The most abundant ions from the survey spectrum, ranging from m/z 350 to 1500, were automatically selected for collision-induced fragmentation using Masslynx. Fragmentation was conducted with argon as collision gas at a pressure of 4 x 10^{-5} bars measured on the quadrupole pressure gauge.

Retention-time *shift of homoserine containing peptides*— Retention-time difference between peptides which were present in two forms, either containing di-aminobutyrate or homoserine was determined by loading 5 µl of sample enriched by COFRADIC as described below on LC QTOF as described above. Resulting tandem mass spectrometry spectra were processed with the MaxEnt3 algorithm embedded in Masslynx, Proteinlynx software to generate peak lists. Generated peak lists were submitted to MASCOT with parameters as described below. Peptide masses determined to be the same peptide with two different forms, i.e. containing either di-aminobutyrate or homoserine were pinpointed in the ion-chromatogrammes and extracted ion chromatogrammes were constructed using Masslynx in order to determine the retention-time difference of both products.

Distribution of TCEP-induced reaction products— To determine relative distribution of TCEP-induced reaction products, 1 nmol of the synthetic azhal-containing peptide Pan016 (PPHHHHHHPPRGFGAzGFR) was reacted overnight at 40 ºC with 10 mM TCEP in 50 mM Hepes pH 8.0 (end concentrations). 20 pmol of reaction products were subsequently separated and identified by LC tandem-MS as described in the above. The relative distribution of reaction products was determined by comparing their relative areas on UV-VIS to the total area of a run of 20 pmol of un-reacted Pan016, the results are presented in Table I.

Cell culture— *E. coli* strain MTD123 (180) was grown aerobically at 37 °C in LB medium. For labelling experiments cells grown overnight in LB medium were transferred to M9 minimal medium containing 6.8 µM $CaCl_2$, 1.0 mM $MgSO_4$, 59.3 µM thiamine·HCl, 57.0 nM Na_2SeO_3, 5.0 µM $CuCl_2$, 10.0 µM CoCl, 5.2 µM H_3BO_3, 99.9 µM $FeCl_3$, 50.5 µM $MnCl_2$, 25.3 µM ZnO, 0.08 µM Na_4MoO_4, 111 mM glucose and 60 mg/l for each of the 19 natural amino acids and 40 mg/l for tyrosine (Sigma-Aldrich, St Louis, USA). Cells were inoculated at OD_{600} 0.1 and allowed to grow into exponential phase before being harvested at OD_{600} 1.0, by spinning down the cells for 10 minutes at 4500 rpm and 4 °C. Cells were then washed twice by resuspending the cell pellet in sterile M9 medium without additives, followed by centrifugation to eliminate traces of methionine. After washing, cells were transferred to M9 minimal medium (see above) in which methionine was replaced by 400 mg/l azhal and cells were allowed to resume growth aerobically at 37 °C. Two biological replicates were labelled in this way and put through the COFRADIC procedure to measure biological variation.

Sample preparation— Unless stated otherwise all the following manipulations in the protocol were carried out in protein low bind tubes (Eppendorf, Hamburg, Germany) to limit losses due to binding to tube surfaces. Labelled

cells were harvested by spinning cells down for 10 minutes at 4500 rpm and 4 °C. Pellets were resuspended in 8 M Urea, 50 mM Hepes pH 8.0 (Sigma Aldrich, St Louis, USA) and lysed by sonication. Lysates were centrifuged for 30 minutes at 15000 rpm and 4 °C to remove cellular debris. Next, samples were dialyzed against 0.5 M Urea, 50 mM Hepes pH 8.0 overnight at 4 °C or against 10 mM of Hepes pH 8.0 for temperature switch samples. Protein content of dialyzed samples was determined using a bicinchoninic acid-based protein assay kit (Pierce, Rockford, USA) following the manufacturers protocol. Samples were then subjected to overnight digestion at 37 °C using a 1:50 (w/w) protease:protein ratio, with trypsin (gold mass spectrometry grade, Promega, Madison, USA). Subsequently, samples were treated with 2 mM TCEP (BioVectra, Charlottetown, Canada) for 5 minutes at room temperature, to reduce disulfide bridges. The duration of this TCEP treatment is too short to induce any reactions with azhal as described above and used to disrupt disulfide bridges only. TCEP treatment was followed by incubation with 5 mM sodium azide and 10 mM iodoacetamide (Sigma Aldrich, St Louis, USA) in the dark at room temperature for 15 min, to oxidize TCEP, and alkylate cysteine residues, respectively. For the primary run of diagonal chromatography 500 µg of the resulting protein digest was loaded onto the SMART system.

Diagonal chromatography— Primary and secondary runs of diagonal chromatography for COFRADIC were carried out with a SMART system (Pharmacia, Uppsala, Sweden) equipped with a 200 µl sample loop and a fraction collector, using a Jupiter Proteo C_{12} column (ID 2 mm, length 150 mm, Phenomenex, Torrance, USA). All solvents used were LC-MS grade (Biosolve, Valkenswaard, The Netherlands). Samples (200 µl) were loaded onto the column using 0.1% TFA in water (Solvent A) at a flow rate of 50 µl/min. for 7 minutes Then the column was washed with this solvent for another 13 min, before a linear gradient to 50% acetonitrile in 0.1% TFA in 75 minutes was applied to elute bound peptides. During the gradient fractions of 1 minute were collected and absorbance of the effluent was continuously monitored at 214, 254 and 280 nm. Fractions collected from 21 minutes until 69 after the gradient start were pooled into 16 pools (A through P, see Figure 1 and Table IV) and lyophilized overnight. Subsequently pools were treated with 10 mM TCEP in 50 mM Hepes pH 8.0 overnight at 40 °C before being reinjected for the secondary run of diagonal chromatography under identical conditions. After the secondary run, fractions which corresponded with the original three primary run fractions, judged on absorbance, plus 3 adjacent fractions on the front and back of an on-diagonal fraction were discarded (Figure 1c). The remaining fractions now termed off-diagonal fractions which contain the shifted reaction products of peptides initially containing azhal were pooled and lyophilized overnight before further analysis.

Mass spectrometric analysis of COFRADIC samples— Off-diagonal pooled fractions were redissolved in 30 µl of 0.1% TFA. For tandem-MS analyses, 5 µl (aerobic growth at 37 ºC) or 10 µl (iTRAQ labelled temperature switch sample spiked with 150 pmol of GluFIB as an internal calibrant) of sample was separated using an Agilent 1100 series LC-system, fitted with a nanoscale reversed-phase high-performance liquid chromatography (RP-HPLC) setup involving Dean switching as described by Meiring *et al.* (231). After loading onto a 2 cm x 100 µm ID C_{18} trapping column (Nanoseparations, Bilthoven, The Netherlands) and washing for 10 minutes at a flow rate of 5 µl/min. with 98% solvent A (0.1% formic acid in water) and 2% solvent B (0.08% formic acid in acetonitrile), the peptides were eluted onto a 63 cm x 50 µm ID C_{18} reversed phase analytical column (Nanoseparations, Bilthoven, The Netherlands) using a linear gradient of 8-30% solvent B for 95 min, at a flow rate of 125 nl/min. The column was interfaced to a QSTAR-XL (Applied Biosystems/MDS Sciex, Toronto, Canada) mass spectrometer, for online electrospray ionization-mass spectrometry (ESI-MS) via a liquid junction with nebulizer using an uncoated fused-silica emitter (New Objective, Cambridge, MA, USA) operating around 4.7 kV (ID, 20 µm, tip ID 10 µm). Survey scans were acquired from m/z 300 to 1,200. The three most intense ions were selected for tandem-MS using automatic selection and dynamic exclusion scripts in Analyst QS 1.1 (max rep = 2; IDA extensions II). Peak lists were generated in Analyst QS 1.1, using the mascot.dll script version 1.6b23, essentially with settings as described on the MASCOT website (http://www.matrixscience.com/help/instruments_analyst.html) with the exception of the precursor mass tolerance for grouping, which was set at 1.0 Da.

Data analysis— Generated peak lists were submitted to the MASCOT search engine 2.1 (Matrix Science, London, UK). The MASCOT search parameters were as follows: cleavage after lysine or arginine unless followed by proline

plus cleavage after methionine, allowing up to 3 missed cleavages, fixed carbamidomethyl cysteine modification and carbamoylation of lysine and the N-terminus as variable modifications. Variable modifications induced by reaction of TCEP with azhal-containing peptides include methionine C-terminally converted to a homoserine lactone after cleaving (analogous to cyanogen bromide cleavage). For the reduction of the azido group a modification was defined on methionine-coded residues as a methionine-residue replaced by di-aminobutyrate ($C_4H_8N_2O$; accurate mass 100.063663) as described previously (143). Besides the reaction products reported earlier, a modification was defined as a methionine-residue replaced by homoserine ($C_4H_7NO_2$; accurate mass 101.047679). Formation of homoserine from azhal has escaped detection in the previous study (143) but was repeatedly observed in the present work. Peptide mass tolerance was set at 0.15 Da and MS/MS tolerance at 0.1 Da. The significance threshold was set to 0.01 resulting in a threshold score of 38. Mudpit scoring and 'require bold red' were applied with an ion score cut-off of 38, in order to have all peptide matches identified at a p-value of <0.01. Mascot performed fragment ion searches with the above settings in a local database of the *E. coli* K12 proteome (4328 proteins; 1381420 residues, release 11 12/07/07, Uniprot consortium, http://beta.uniprot.org/). To estimate false positive rates in protein identification we also performed fragment ion searches against a decoy database, which was a shuffled version of the *E. coli* K12 proteome made using the Peakhardt decoy database builder (Medizinisches Proteom Center, Bochum, Germany; http://www.medizinisches-proteom-center.de). False positive rates were estimated by dividing the total number of protein hits from the decoy database by the total number of protein hits from the *E. coli* K12 database times 100 percent.

Peptide samples subjected to LC tandem-MS for protein identification with MASCOT contained both TCEP-induced-reaction products from azhal-labelled peptides and unmodified peptides not derived from azhal-containing peptides by TCEP treatment. To remove the latter species, we selected manually in MASCOT only those peptides with the variable modifications homoserine lactone, di-aminobutyrate and homoserine and the unmodified C-terminal peptides which resulted from cleavage after a methionine residue. Next the MASCOT search was performed again with this selection using the same settings as described above, in order to recalculate MASCOT protein scores and protein coverage based solely on reaction products. The resulting proteins, representing newly synthesized proteins made during the labelling time with azhal, were exported as a csv-file for further analysis.

Determination of chromatographic separation of peptides by COFRADIC — The exported mgf-file of each LC-MS run of 16 runs was submitted to MASCOT separately as described in the above. The results of each search were exported as a csv-file and imported into Excel, where reaction products were selected using ASAP-utilities add on for Excel (A Must In Every Office B.V., Zwolle, The Netherlands, http://www.asap-utilities.com/) conditional select on the different reaction products and cleavage after methionine. For each pool duplicate identifications were filtered out, after which all runs were combined and the number of runs in which each reaction product was identified was counted using a pivot table. The results are shown in Tables I and II.

Annotation of proteins to GO terms— To annotate the identified (newly synthesized) proteins with Gene Ontology terms, the list of proteins was assigned using GO-miner (232), run locally with a Derby database engine using the Uniprot database and the organism set at: 562 (*E. coli*) and evidence codes at: 'all'. Resulting GO-annotation categories and the corresponding proteins were exported to Excel. With the use of the generic GO-SLIM set (Gene Ontology Consortium, http://www.geneontology.org/), from which strictly eukaryotic terms were removed, the proteins detected were assigned to parent GO categories to look at the distribution of the newly synthesized proteins across the various different biological processes and cellular localizations. To asses the relative over- or under-representation of mapped proteins per category, the number of mapped proteins per category was divided by the sum of the mapped proteins of all the categories in the table times 100 percent. This was done for both categories representing biological processes and cellular localizations as well as for both the entire proteome and the newly synthesized proteins dataset. Next, the relative ratio of representation was calculated for the categories by dividing the percentage per category for the dataset by the percentage per category of the proteome. Thus, ratios greater than 1 account for relative over-representation of mapped proteins in the category of the newly synthesized proteins compared to the proteome and ratios smaller relative under-representation.

4

Immediate changes in both protein levels and newly synthesized proteins following a change in growth temperature in *E. coli*

Summary

Pulse-labelling with the methionine analogue azhal has been used for the identification of hundreds of newly synthesized proteins in *E. coli* on a short time-scale. However a quantitative mass-spectrometric approach is essential to enable research into how different environmental conditions affect new protein formation in *E. coli*. We applied azhal pulse-labelling in combination with iTRAQ for quantitation and COFRADIC isolation of labelled peptides to determine changes in newly synthesized proteins immediately following a heat shock in *E. coli*. In addition by extension of our analytical strategy to determine changes in total protein levels on the same time-scale stable or labile proteins can be identified.

Measurement of the relative amounts of 344 proteins newly synthesized in 15 minutes upon a switch of growth temperature from 37 ºC to 44 ºC showed that nearly 20% in- or decreased more than two-fold. Amongst the most up-regulated proteins many were chaperones and proteases, in accordance with the cells response to unfolded proteins due to heat stress. While collation of changes in protein levels with changes in newly synthesized proteins showed that the vast majority of proteins were stable, only a subset of 5 proteins (PspA, IbpB, σ^{32}, AhpC, and CysK) were found to have a higher turnover rate. Finally, comparison of our data with results from previous microarray experiments revealed the importance of regulation of gene expression at the level of transcription of the most elevated proteins under heat shock conditions and enabled identification of several candidate genes whose expression may predominantly be regulated at the level of translation. This work demonstrates for the first time the use of a bioorthogonal amino acid for proteome-wide detection of changes in the amounts of proteins synthesized during a brief period upon variations in cellular growth conditions.

Introduction

Changes in protein levels in cells during adaptation from one environmental condition to another may be regulated both by transcription and translation. However, little is known about the contribution of translational regulation. This requires information about genome-wide changes in translation rates and mRNA levels upon environmental perturbations. Measuring protein synthesis and degradation-rates on a proteomic scale is an important step towards modelling the kinetics in complicated cellular response networks. Pulse-labelling with the methionine analogue azhal enables proteome-wide identification of proteins that are newly formed during the pulse-labelling period, while different enrichment schemes can facilitate short pulse-labelling times, increasing temporal resolution over stable-isotope based pulse-labelling approaches (see *Chapter 1 and 3*). In *Chapter 3* we demonstrated how a selective reaction against the azide-moiety of azhal in combination with an enrichment scheme based on an induced retention-time shift enables a short pulse-labelling time window of only 15 minutes in *E. coli*. However to enable investigations into how different environmental conditions affect the formation of new proteins on a proteome-wide scale, a quantitative mass spectrometric approach is essential to compare differences in the synthesis of new proteins between different samples. In this chapter we show that through combining azhal pulse-labelling with relative quantitation by means of iTRAQ labelling (183) to compare different samples, relative changes in newly synthesized proteins after a 15 minute pulse-labelling period can be measured.

In addition we show that the approach is easily extended to also determine changes in total protein levels on the same time-scale as changes in newly synthesized proteins. Total protein levels are determined by reporter ions from peptides that do not contain azhal or methionine, as these represent the sum amount of newly synthesized and pre-existing material after the labelling period. Comparison of relative changes in total protein levels to those in newly synthesized proteins allows identification of stable and labile proteins for those proteins that undergo significant regulation upon heat shock. Information on the stability of proteins is an interesting extension. Although there is data that the majority of abundant proteins in *E. coli* cells is relatively stable (109) and has a half-life of at least a few hours (110, 111), while a small pool is rapidly degraded (90), no proteome-wide information on protein turnover for individual proteins is available.

To study relative changes in newly synthesized proteins labelled with azhal and the effects on total protein levels, we chose the heat shock response in *E. coli,* as it has been studied in detail before it therefore seems a good system to validate our approach (48, 103-105). Heat shock in *E. coli* is defined as the cellular response to an increase in growth temperature, and is accompanied by the upregulation of a defined set of 'heat shock' proteins. The increase of heat shock proteins is controlled by *rpoH* which encodes the heat shock transcription sigma factor σ^{32} (226-228, 233). Nearly 100 genes have been identified to be part of the σ^{32}-regulon (229, 230), of which some 18 genes encode chaperones and proteases. The upregulation of both chaperones and proteases seems to be aimed at restoring impaired

protein folding at higher temperature and to degrade misfolded proteins. The intracellular concentration of σ^{32} shows a rapid transient increase upon heat shock followed by a decrease to reach a new steady state within 10-15 minutes (50, 106, 107). This affects transcription of σ^{32} regulated genes concomitantly and also induces expression of heat shock proteins within this time-frame (106, 107). In addition the transcription of a large number of other genes is rapidly up- or down-regulated upon heat shock (234, 235).

The early changes in protein synthesis, upon the change in growth temperature, were examined and quantitative data is presented for 344 newly synthesized proteins. In addition the changes in total protein levels during the initial 15 minutes after the temperature switch were quantified for 292 proteins. Comparison of the changes in newly synthesized proteins with relative mRNA levels enables assessment of the separate contributions of transcription and translation to the regulation of gene expression and allows identification of candidate genes that could be subject to post-transcriptional regulation. This demonstrates for the first time a proteome-wide, bioorthogonal approach for relative quantitation of proteins synthesized during a small time-window, upon a change in growth conditions.

Results

Quantitation of newly synthesized proteins induced by heat shock— To study the immediate changes in newly synthesized proteins following heat shock, quantitative azhal pulse-labelling to measure changes in the amount of newly synthesized protein after an up-shift in growth temperature from 37 ºC to 44 ºC. Two cultures (biological replicates) were grown into exponential phase before being harvested and washed prior to pulse-labelling, subsequently cells were pulse-labelled in minimal medium containing azhal for 15 minutes at either 37 ºC or 44 ºC. We used iTRAQ (183) to detect changes in the amounts of proteins synthesized during this brief period by quantitative mass spectrometry. After the 15 minute pulse with azhal at either 37 °C or 44 °C, cells were harvested and proteins were extracted and digested as described above. The different digests were treated with different iTRAQ labels and then mixed in equivalent amounts. Newly synthesized proteins were identified by azhal-labelled peptides only, enriched by COFRADIC.

Because only methionine containing peptides can be labelled with azhal and used for the identification of newly synthesized proteins, the number of available peptides per protein that can be used for identification and quantitation sharply declines as shown in the supplemental data using an in-silico digest (Supplemental Figure 1). Consequently the contribution of single peptide protein identifications increases, as compared to the standard approach when all tryptic peptides can be used for identification (Supplemental Table I). However, a large part of the *E. coli* proteome (93.1%) is still predicted to be represented by this subset of peptides (Supplemental Figure 1) and proteins identified do not seem to be biased towards higher methionine content (Supplemental Table I).

A total of 394 newly synthesized proteins were identified after the 15 minute pulse. Of the identified newly synthesized proteins, 344 could be quantified using the iTRAQ reporter

ions, according to the criteria formulated in the experimental procedures. Upon an elevation in temperature the relative abundance of 64 newly synthesized proteins significantly ($p< 0.05$) in- or decreased more than a factor of two, (1 on ^{2}log scale) while 65 changed significantly ($p< 0.05$) by only a factor of 1.5-2 (0.58-1 on ^{2}log scale). The relative abundance of the remaining 216 newly synthesized proteins changed less than by a factor of 1.5 or did not change significantly at all during the 15 minute period after temperature switch (Figure 1).

Expression of heat shock-related proteins is induced by the *rpoH* gene product σ^{32}. This alternate sigma-factor is induced by heat shock and growth at higher temperatures during exponential aerobic growth (226-228). The level of newly formed σ^{32} was found to be up-regulated more than twofold within the first 15 minutes after the temperature change from 37 °C to 44 °C (Figure 1). Large scale transcriptomics studies with respect to σ^{32}-inducible genes have been carried out in which 97 genes were identified to be part of the σ^{32}-regulon (229, 230). From this σ^{32}-regulon, 28 gene products were identified, among these a number have chaperone functions, and aid in refolding proteins that are misfolded due to the temperature increase. The levels of newly synthesized proteins increased dramatically for HtpG, ClpB, IbpB, GroEL (CH60), DnaK and GrpE immediately following the temperature increase. Together with chaperones that aid in (re)folding of proteins at higher temperature, proteases are another important class of proteins induced during heat shock. Proteases degrade misfolded proteins and aid chaperones in (re)folding proteins during growth at higher temperatures. From the 28 heat shock-inducible proteins identified that are part of the σ^{32}-regulon, eight have protease functions. Of these four were found to be up-regulated in their respective levels of new protein formation more than two-fold (DegP, HslU, HtpX, LoN), while newly synthesized protein levels of the four others (ClpP/ClpA, FtsH and HflK) did not change when growth temperature increases (Figure 1).

Apart from the heat shock proteins that are part of the σ^{32}-regulon, we also found that levels of newly synthesized PspA increased significantly (3.3 fold ^{2}log scale). PspA is part of the phage shock PspABCDE regulon, which is induced by filamentous phage infection, and various environmental stresses including severe heat shock (236-239). Induction of PspA by increased temperature is independent of σ^{32} but is mediated by σ^{54}. However σ^{32}-mutants have a prolonged increase of PspA levels following heat shock (236, 237, 240), probably due to the lack of a proper heat shock response, suggesting cross-talk between the two pathways. Finally there is also a large group of proteins which have significantly lower levels of newly synthesized proteins after the increase in temperature. This is a group containing proteins with diverse functions, such as ribosomal proteins, cysteine biosynthesis/sulphur metabolism.

Among our dataset of 344 proteins, we could identify 15 species for which relative synthesis-rates under heat shock conditions have been measured previously (50, 106, 107, 122, 241). In all cases our data are remarkably similar with this previous work The relative amounts of three proteins of which the relative synthesis-rates increased dramatically during heat shock, namely GroEL (CH60) (122, 241), chaperone protein ClpB (106, 122) and σ^{32} (RP32) (50, 106, 107) where also highly increased under our assay conditions. In general

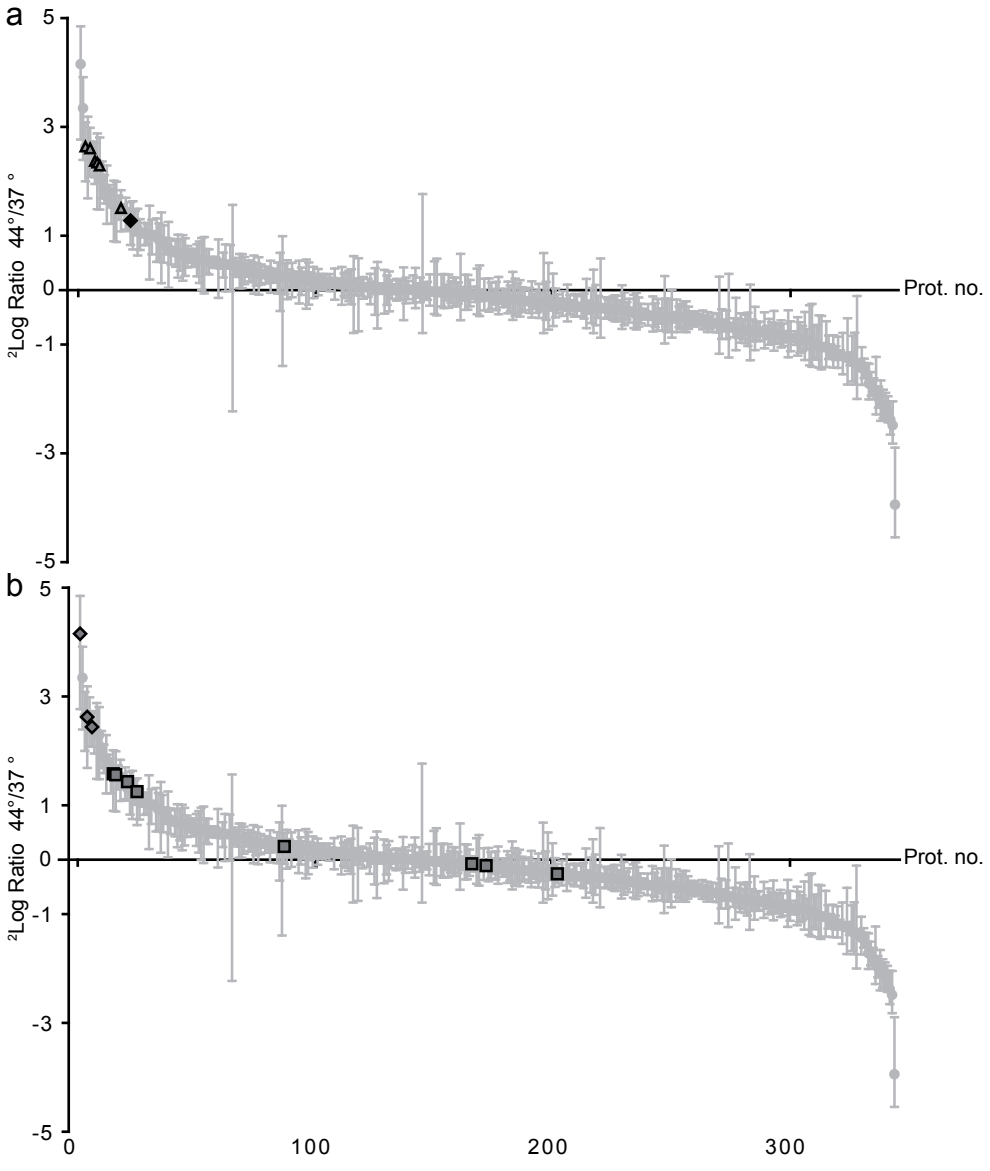


FIGURE 1. **Relative abundance of newly synthesized proteins 15 minutes after an increase in growth temperature**. Plot of relative abundance of newly synthesized proteins of cells grown at 44 °C compared to those grown at 37 °C during 15 minute pulse-labelling period with azhal. In panel (**a**) the increase in newly synthesized heat shock related proteins with a chaperone function (dark grey triangles), HtpG, ClpB, DnaK, IbpB, GroEL (CH60) and GrpE and the heat shock sigma factor, σ^{32} (black diamond) are shown. In panel (**b**) heat shock proteins with protease function (dark grey squares) DegP, HslU, HtpX, LoN, FtsH, ClpA, HflK, ClpP and three proteins (UxuA, UxuB, UxaC) involved in hexuronide and hexuronate degradation pathway (dark grey diamonds) are shown. Proteins are ordered from most to least increased in new synthesis after an increase in growth temperature, error bars denote standard deviation of peptide ratios obtained from biological replicates.

the relative amounts of newly synthesized species of the remaining 12 proteins, all involved in protein biosynthesis, were slightly decreased or not changed at all, in agreement with the slightly decreased synthesis-rates measured previously under heat shock conditions (122). These results strongly indicate that pulse-labelling with azhal is a reliable method to detect changes in the amounts of protein synthesized in a brief time frame upon changes in growth conditions.

Quantitation of immediate changes in total protein levels in response to heat shock— To estimate how the changes in newly synthesized proteins affect total protein levels on the same time-scale we also measured peptides that do not contain methionine or azhal. Just as azhal labelled peptides represent exclusively newly formed proteins, peptides that do not contain azhal or methionine represent the total protein content as they are made up of both pre-existing and newly synthesized proteins. To measure these peptides we analyzed the 'un-shifted' fractions obtained during COFRADIC by LC tandem-MS. Peptides that do not contain azhal or methionine are found in off-diagonal fractions as well, their presence is due to tailing of main chromatographic peaks that contain the bulk of unlabeled non-shifting peptides. The combined database searches from data acquired from shifted and non-shifted pooled fractions yielded a total of 1060 peptides identifying 435 proteins of which 292 could be quantified by the iTRAQ reporter ions. Upon an elevation in temperature the total protein levels of 8 proteins significantly ($p< 0.05$) in- or decreased more than a factor of two, (1 on ^{2}log scale) while 5 changed significantly ($p< 0.05$) by only a factor of 1.5-2 (0.58-1 on ^{2}log scale). The relative abundance of the remaining 279 proteins changed less than by a factor of 1.5 or did not change significantly at all during the 15 minute period after temperature switch (Figure 2).

In accordance with the elevated newly synthesized proteins, the proteins which have increased levels following the elevation of growth temperature mostly carry out chaperone and protease functions and are part of the σ^{32}-regulon. The heat shock sigma-factor increased its levels more than two-fold in the culture grown at 44 °C during pulse-labelling, in accordance with earlier observations (50, 106, 107). Concomitantly most of the chaperones IbpB, IbpA, ClpB, HtpG, DnaK, GroES (CH10), GroEL (CH60) and GrpE under its transcriptional control that were detected had increased levels from 1.4 to almost seven-fold (Figure 2). In contrast only one of the proteases had significantly changed total protein levels (DegP), while four others (HslU, ClpA, ClpX and FtsH) did not significantly change within the first 15 minutes upon heat shock. Total levels of PspA also increased significantly upon heat shock in accordance with it the increased levels of newly synthesized proteins described in the above. Conversely there were only two proteins (SyW and CysK) that had significantly decreased protein levels 15 minutes after the increase in growth temperature.

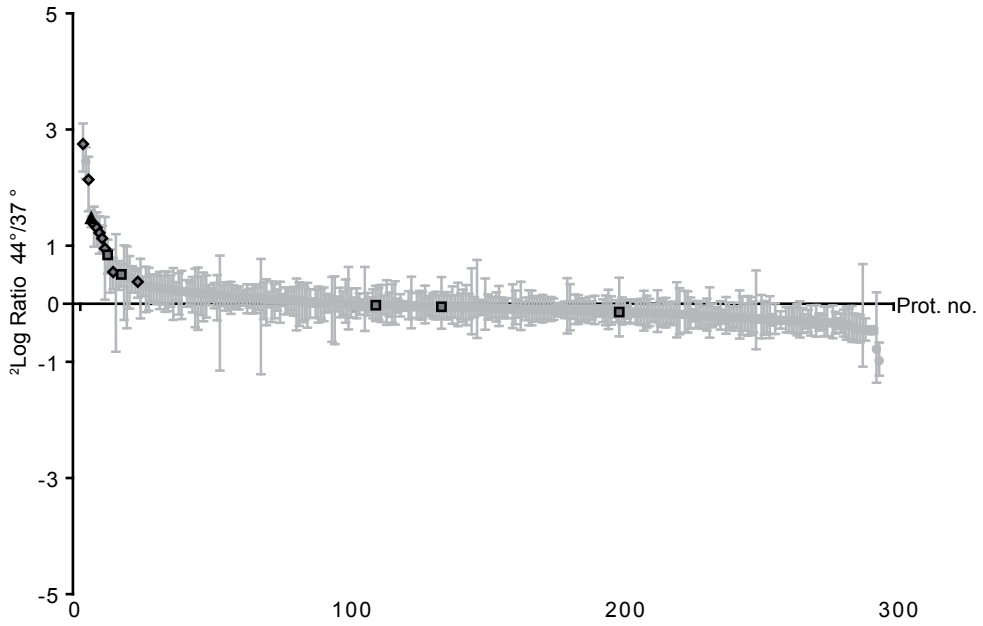


FIGURE 2. **Changes in total protein levels 15 minutes after an increase in growth temperature**. Total protein levels are shown for the heat shock sigma factor, σ^{32} (black triangle), heat shock proteins with chaperone function: IbpB, IbpA, CpbB, HtpG, DnaK, GroES (CH10), GroEL (CH60), GrpE and DnaJ (dark grey diamonds) or protease function: DegP, HslU, ClpA, ClpX and FtsH (dark grey squares). Proteins are ordered from most to least increased in total levels after an increase in growth temperature, error bars denote standard deviation of peptide ratios obtained from biological replicates.

Data on both changes in total protein levels and newly synthesized proteins during pulse-labelling allows identification of labile and stable proteins—By measurement of both changes in protein levels and newly synthesized proteins through the iTRAQ reporter ions stemming from azhal/methionine lacking- and azhal-containing peptide populations, our dataset contains 176 identified proteins for which both changes in total protein levels as well as changes in newly synthesized proteins were determined. In Figure 3 the data are ordered according to the relative levels of newly synthesized proteins (dots), while corresponding total levels are represented by diamonds at the same position along the X-axis. Overall, it is clear that the protein levels change to a lesser extent than the newly synthesized proteins newly synthesized proteins during the first 15 minutes after the change in growth temperature. However, we also identified five proteins, PspA, IbpB, σ^{32}, AhpC, and CysK, of which the relative total levels changed significantly upon the change in growth temperature, practically as much as the relative levels of the newly synthesized species (Figure 3a). Apparently, almost all pre-existing protein molecules of this group have been replaced by newly synthesized polypeptides during the pulse. So, the half-life of these proteins lies well within the pulse-labelling time used. For these proteins synthesis and degradation-rates are large compared to their total cellular levels during exponential growth at 37 °C. As a consequence a change in new protein formation will also affect the total cellular levels concomitantly

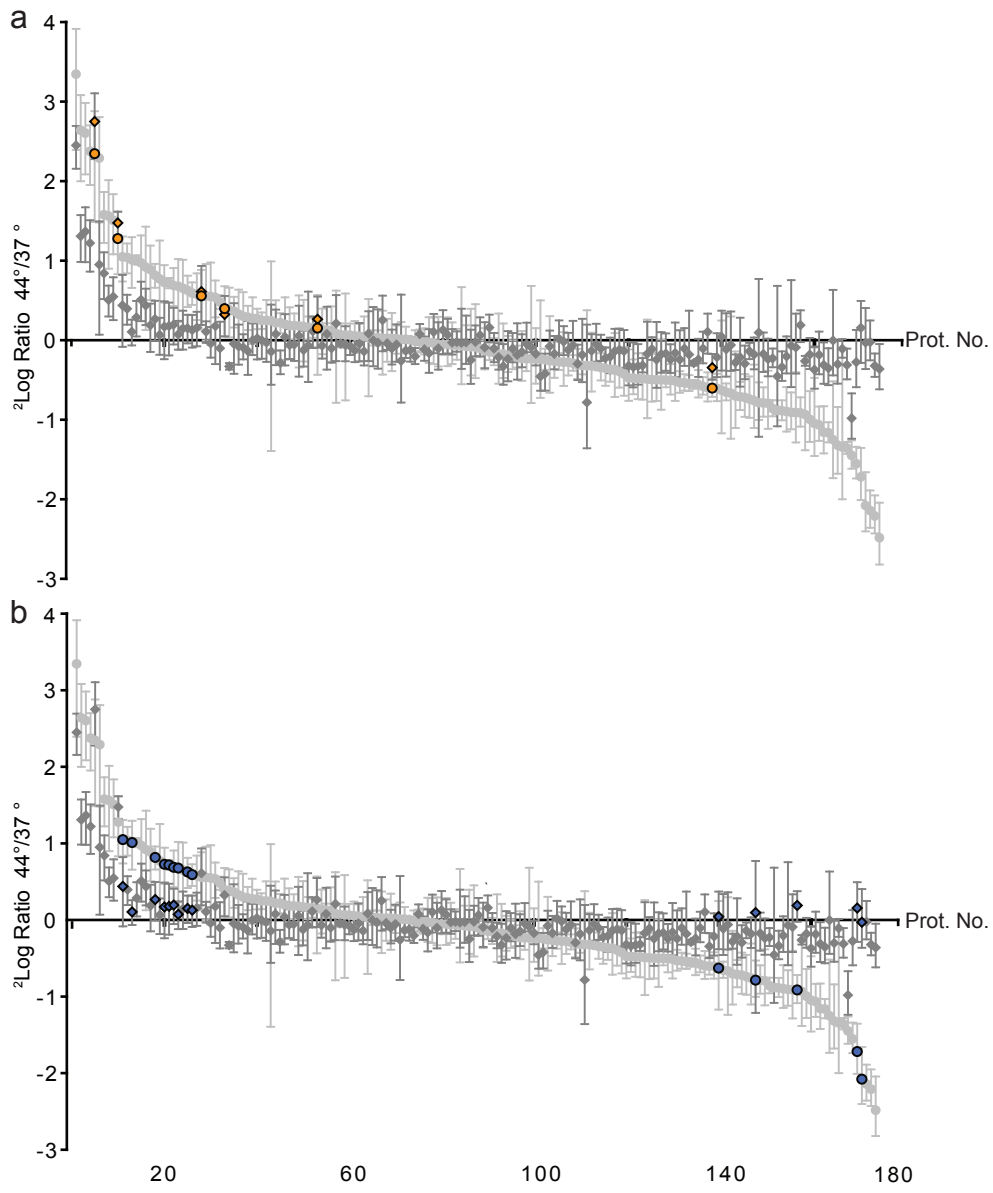


FIGURE 3. **Changes in protein synthesis compared to changes in protein levels upon onset of a change in growth temperature.** Light grey or coloured dots, relative change of newly synthesized proteins; dark grey or coloured diamonds at the same position on the X-axis as a dot are relative total protein levels of the same protein 15 minutes after the switch from 37 to 44 °C. Proteins are ordered from most induced to most repressed synthesis. In panel (**a**) proteins having a high turnover rate are marked, namely IbpB , σ^{32}, AhpC, AhpF and KatG (orange dots and diamonds). Coloured symbols in panel (**b**) represent proteins with a low turnover rate. Blue dots and diamonds, ribosomal proteins.

on the short time-scale used for pulse-labelling. The high turnover of σ^{32} has been reported before; it has a half-life of less than a minute during growth at normal temperatures, and is transiently stabilized during temperature increase (106, 107). This is thought to occur due to the an increase of unfolded protein as alternative substrate for FtsH, which decreases the degradation of σ^{32} (105). IbpB, a member of the family of small heat shock proteins, also has been found to have a high turnover rate (242). Surprisingly enough PspA on the other hand has been reported to be a stable protein (237), however the chase-experiment was conducted at 37 °C following induction of PspA synthesis by a temperature increase from 37 °C to 50 °C during pulse-labelling. So it is possible that following heat shock PspA degradation-rate is decreased in analogy to σ^{32}. If the half-life under normal growth conditions is short, then even though it is stabilized during heat shock, the increase in protein levels will mirror the increase in newly synthesized proteins as is the case for σ^{32}. No information on turnover was available for the other two proteins.

In contrast to the proteins for which the relative total levels change to almost the same degree as the relative newly synthesized levels upon heat shock, a significant difference between relative total levels and relative levels of newly synthesized material after a pulse of 15 minutes was noted for the majority of proteins. This is shown in Figure 3b by blue symbols for ribosomal proteins, but also the case for various chaperones and proteases. The difference between levels of total and newly synthesized protein can be explained by assuming that these proteins have a half-life that exceeds the pulse-labelling time used. Consequently, a large change in synthesis will not affect the total protein levels to the same extent in the short time frame used for pulse-labelling; only a continued altered rate of new formation will eventually change the protein levels to the same extent.

The observation that the protein levels for the majority of the proteins does not change significantly during the 15 minutes of elevated growth temperature also for most stable proteins for which protein new formation decreases, suggests that pre-existing proteins are mostly stable and not affected by the increased temperature. This is illustrated by the relative levels of methionine containing peptides in Figure 4. Just as azhal labelled peptides represent newly formed proteins, and non-azhal/methionine containing peptides the total protein levels, the iTRAQ reporter ions from these methionine containing peptides represent the relative levels of pre-existing proteins at 37 °C and 44 °C. As is clear from Figure 4 the relative levels of pre-existing proteins at 44 °C hardly change relative to those at 37 °C during the 15 minutes of pulse-labelling for the 87 proteins for which methionine containing peptides were found. This means that the elevated temperature does not change the half-life of the pre-existing proteins to such an extent that it is appreciable on the 15 minute period used for pulse-labelling. Apparently newly synthesized proteins are affected most with respect to stability, as these still need to be folded, while already folded proteins are not denatured and degraded at the temperature increase used here.

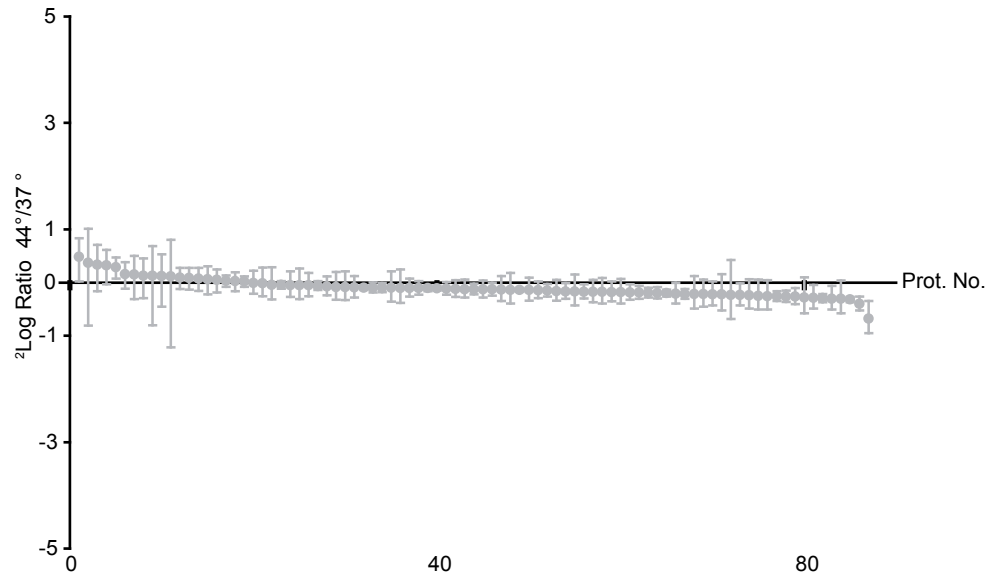


FIGURE 4. **Relative stability of pre-existing proteins during 15 minutes after an increase in growth temperature.** Plot of relative abundance of pre-existing proteins in cells grown at 44 °C compared to those grown at 37 °C during 15 minute pulse-labelling period with azhal. Proteins are ordered from most to least increased after an increase in growth temperature, error bars denote standard deviation of peptide ratios obtained from biological replicates. No major differences in the relative abundance of pre-existing proteins is obvious, which suggest that on the time-scale used for pulse-labelling, the degradation of pre-existing proteins does not play an important role for those proteins which were measured.

Comparison of changes in newly synthesized proteins and transcript levels suggest transcriptional regulation upon heat shock—Among the 24 proteins of which the relative amounts synthesized during the pulse were increased by at least a factor 2.4 (1.3 on ^{2}log scale), 14 belonged to the group of σ^{32}-regulated heat shock proteins of which the corresponding transcripts increased considerably during heat shock as measured by Richmond *et al.* (235) and Harcum *et al.* (234) (Table I). Remarkable is the presence of three proteins (UxuA, UxaC and UxuB), involved in hexuronide and hexuronate catabolism in the top six of proteins, of which the amounts synthesized during the pulse of azhal is increased, along with a corresponding increase in transcript level. Of the remaining seven proteins in the top 24, the corresponding mRNA levels are likewise increased in six cases and not measured in one case. The strong correlation between our proteomic data and transcriptomic data available in literature (234, 235), although measured at different temperature shifts, underscores the importance of regulation of gene expression at the level of transcription of the most elevated proteins under heat shock conditions.

Interestingly, the strong correlation between transcript levels and levels of newly synthesized proteins in the group of most up-regulated proteins did not exist in the group of proteins of which the amounts synthesized during the azhal pulse was decreased by a factor of 2.4 or more (Table II).

TABLE I
Highly up-regulated proteins upon heat shock

Gene name	Protein	Protein ratio†	Transcript ratio‡	Transcript ratio‡‡
uxuA	Mannonate dehydratase	3.8	7.6	1.29
pspA	Phage shock protein A	3.3	28.2	1.82
htpG	Chaperone protein htpG	2.6	33.8	4.26
uxaC	Uronate isomerase	2.6	1.8	0.77
clpB	Chaperone protein clpB	2.6	36.5	4.32
uxuB	D-mannonate oxidoreductase	2.4	3.2	0.47
dnaK	Chaperone protein dnaK	2.4	58.5	3.05
ibpB	Small heat shock protein ibpB	2.3	327.5	5.73
mopA	60 kDa chaperonin GroEL	2.3	37.9	3.83
ybeD	UPF0250 protein ybeD	2.1	1.6	2.93
relB	Antitoxin RelB	1.9	3.7	1.47
hdhA	7-alpha-hydroxysteroid dehydrogenase	1.9	2.5	0.17
ybdQ	Universal stress protein G	1.8	1.6	2.78
htrA	Protease do	1.6	9.6	3.11
hslU	ATP-dependent hsl protease ATP-binding subunit hslU	1.6	10.3	2.50
yfiA	Ribosome-associated inhibitor A	1.5	2.3	1.77
sdaA	L-serine dehydratase 1	1.5	23.6	3.56
grpE	Protein grpE	1.5	24.1	4.13
yibT	Uncharacterized protein yibT	1.5	n.d.	n.d.
htpX	Probable protease htpX	1.4	36.1	5.34
recN	DNA repair protein recN	1.3	1.7	1.08
rpoH	RNA polymerase sigma-32 factor	1.3	4	2.33
trxC	Thioredoxin-2	1.3	2.4	1.43
lon	ATP-dependent protease La	1.3	20.3	3.51
rplD	50S ribosomal protein L4	1.0	- 4.1	-1.27
yfgB	Ribosomal RNA large subunit methyltransferase N	1.0	- 1.2	-0.64
gltA	Citrate synthase	1.0	- 3	-1.52
rpsH	30S ribosomal protein S8	1.0	- 1.4	-1.93

† Relative protein ratio 44 ºC/37 ºC of proteins synthesized during 15 minute labelling period upon a change in growth temperature. ‡ Relative transcript ratio 50 ºC/37 ºC as reported by Richmond *et al.* (235), ‡‡ Relative transcript ratio 50 ºC/37 ºC as reported by Harcum *et al.* (234) n.d. transcript ratio not determined. (^{2}logscale)

Of the 19 proteins in this group, 7 are reported to have increased transcript levels upon an elevation in growth temperature by Richmond *et al* (235), while in the study of Harcum *et al.* (234) 6 of these are actually reported to be significantly down-regulated under the same conditions. There were 3 proteins (NlpA, YfdZ and SerA) which had significant decreased levels of newly synthesized proteins, while transcript hardly changed upon an increase in temperature reported by both transcriptomic-studies (234, 235) (Table II). In addition four proteins were identified of which the synthesis increases ~2-fold (1 fold on ^{2}log scale), whereas transcript levels change in the opposite direction. This suggests regulation at a post-transcriptional level for these proteins (Table I).

TABLE II
Highly down- regulated proteins upon heat shock

Gene name	Protein	Protein ratio†	Transcript ratio‡	Transcript ratio‡‡
cysM	Cysteine synthase	-1.3	- 1.6	-0.9
purA	Adenylosuccinate synthetase	-1.3	- 4.4	-1.2
b1680	Cysteine desulfurase	-1.3	2.5	-4.1
rho	Transcription termination factor rho	-1.3	- 7.3	-1.4
serA	D-3-phosphoglycerate dehydrogenase	-1.4	n.d.	0.2
tyrB	Aromatic-amino acid aminotransferase	-1.4	- 2.4	-1.1
cysK	Cysteine synthase A	-1.4	- 3.8	-3.3
yihK	GTP-binding protein typA/BipA	-1.6	- 4.6	-0.7
rplN	50S ribosomal protein L14	-1.7	- 1.7	-2.3
cysN	Sulfate adenylyltransferase subunit 1	-1.7	2.3	-3.5
gcd	Quinoprotein glucose dehydrogenase	-1.8	2.2	-0.9
cysJ	Sulfite reductase flavoprotein alpha-component	-1.9	1.3	-3.1
nlpA	Lipoprotein 28	-2.0	0	-0.1
rplY	50S ribosomal protein L25	-2.1	0.1	-1.7
cysD	Sulfate adenylyltransferase subunit 2	-2.1	2.9	-3.6
ydfY	Protein gnsB	-2.1	- 1.2	0.6
hlpa	Chaperone protein skp	-2.2	- 4.8	-1.3
b2379	Uncharacterized aminotransferase yfdZ	-2.4	1.3	-0.1
oppA	Periplasmic oligopeptide-binding protein	-2.5	- 4.6	-3.2
cysP	Thiosulfate-binding protein	-3.9	1.9	-2.4

† Relative protein ratio 44 ºC/37 ºC of proteins synthesized during 15 minute labelling period upon a change in growth temperature. ‡ Relative transcript ratio 50 ºC/37 ºC as reported by Richmond *et al.* (235), ‡‡ Relative transcript ratio 50 ºC/37 ºC as reported by Harcum *et al.* (234) n.d. transcript ratio not determined. (^{2}logscale)

Discussion

We used pulse-labelling with azhal to determine the relative abundance of proteins synthesized during the transition period after a change in growth temperature from 37 to 44 ºC. By the use of iTRAQ for relative quantitation of newly synthesized proteins, the relative abundance of 344 newly synthesized proteins was determined upon a change in growth temperature. Amongst proteins highly up-regulated after the temperature switch there were many heat shock inducible chaperones and proteases, all part of the σ^{32}-regulon. In addition, PspA which has been previously reported to be induced after heat shock independent of σ^{32} (237) was also identified amongst the most up-regulated newly synthesized proteins after the temperature switch. Proteins, for which radiolabelling data was available, showed high similarity to levels of newly synthesized proteins determined by azhal-labelling. This further validated that the protein expression measured here is due to temperature-induced changes in gene expression, and not the result of labelling cells with azhal.

It should be noted however that there may already be a significant contribution of degradation to the amount of each protein formed in 15 minutes. This is dependent on the particular protein's half-life. We also determined the changes in total levels on the same time-scale of pulse-labelling by iTRAQ reporter ions stemming from non-azhal/methionine containing peptides. Comparison of total protein levels and newly synthesized proteins for those proteins that showed significant changes upon heat shock revealed two distinct groups of proteins. The first group comprised of five proteins (PspA, IbpB, σ^{32}, AhpC, and CysK) had protein levels that changed as much as the levels of newly synthesized proteins upon heat shock. This can be explained by assuming these are labile proteins that have a short half-life within or not much longer than the pulse-labelling time used. This means that a change in new protein formation, which can be the result of a change in protein synthesis-rate, protein degradation-rate or a combination of the two, will also significantly affect protein levels on the short time-scale used for pulse-labelling.

The large majority of proteins that were significantly regulated upon heat shock showed a marked discrepancy between changes in protein levels and newly synthesized proteins. In contrast to the group of labile proteins mentioned in the above, these are stable proteins with a half-life that exceeds the pulse-labelling time used, as a change in new protein formation does not affect the total protein levels on a short time-scale to the same degree. Among these stable proteins were chaperones and proteases of the σ^{32}-regulon and the ribosomal proteins. The change in newly synthesized protein levels for these proteins could be the result of an increase or decrease of protein synthesis, as protein half-life should not play a role on the short-time-scale of pulse-chase labelling. This is demonstrated for most of the pre-existing proteins, which did not significantly change their levels within the pulse-labelling time with respect to the cells grown at 37 °C. However, newly formed proteins are probably more affected by the increase in temperature than pre-existing ones as these still have to be folded, and are thus more likely to be degraded if they misfold in spite of chaperone to aid in this process. Consequently the relative levels found at these two different

temperatures probably do not directly reflect the relative synthesis-rates for each protein found, this depends on whether the relative stability of the newly synthesized polypeptides are also similar under the different temperatures.

Comparison of the results with transcript data from literature revealed that many proteins that are highly up-regulated upon a temperature switch seem to be regulated at the transcriptional level, as increased transcript levels corresponded with an increase of protein synthesis, as determined by azhal incorporation. Examples to illustrate this are the genes *uxuA*, *uxuB* and *uxaC* that encode enzymes that catalyze different steps in the catabolism of hexuronides and hexuronates to 2-keto-3-deoxy-gluconate (KDG). KDG is metabolized by the Entner-Doudoroff pathway and enters the lower part of glycolysis (243). There are no previous reports about temperature induction at the protein level of members of this pathway. However UxuA, UxuB and UxaC were found to be highly up-regulated upon heat shock in this study, in good correlation with altered transcript levels reported before (235). A possible explanation for the measured increase in transcription could be instability of the transcriptional repressors of these genes (ExuR, UxuR) at higher temperatures.

In contrast to the highly up-regulated heat shock proteins, there was poor correlation between transcript data and relative levels of newly synthesized proteins in a group of three highly down-regulated proteins. Increased turnover of these proteins at higher temperature offers one explanation. While for a group of six other proteins that were highly down-regulated with respect to newly synthesized proteins, transcriptomic studies contradicted each other, whether proteins were up or down-regulated upon heat shock. For a group of four proteins found to be up-regulated, with transcript levels going down significantly, upon elevation of the growth temperature, post-transcriptional regulation is a possible explanation. This shows that azhal-labelling is very suitable to identify candidates that may be subject to post-transcriptional regulation.

The first proteome-wide approach to quantitation of newly synthesized proteins by azhal-labelling is demonstrated here using COFRADIC. Altogether, the fact that no severe azhal-related disturbances were obvious, combined with the large number of newly synthesized proteins identified and quantified, makes azido-peptide isolation by COFRADIC in combination with iTRAQ an excellent tool for both identification and quantitation of transient changes in protein expression. While the measurement of total protein levels on the same time-scale allows identification of stable and labile proteins among those proteins that are significantly regulated by the environmental stress. The pulse-labelling technique described is uniquely suited to follow an adaptation to changes in the environment of *E. coli.* Furthermore comparison with transcript data allows for screening for different types of regulation in response to a change in environment. The presented method can open up new avenues in systems biology research, by filling the gap of information between transcriptomics and proteomics and allow for new input into advanced modelling of cellular networks.

Supplemental Data—Supplemental Figures and Tables can be found in the addendum section on pages 136-137. Protein identification and quantitation data can be found online at http://www.mcponline.org as supplemental information for references (141) and (169).

Experimental Procedures

Synthesis of L-azhal— L-azhal was synthesized from L-Boc-2,4-di-aminobutyric acid (L-Boc-DAB, Chem-Impex, Wood Dale, USA) by diazotransfer (198) using Triflic azide (TfN_3) as previously described (ref).

Cell culture— The methionine-auxotrophic *E. coli* strain MTD123 (180) was grown in M9 minimal medium containing 6.8 µM $CaCl_2$, 1.0 mM $MgSO_4$, 59.3 µM thiamine·HCl, 57.0 nM Na_2SeO_3, 5.0 µM $CuCl_2$, 10.0 µM CoCl, 5.2 µM H_3BO_3, 99.9 µM $FeCl_3$, 50.5 µM $MnCl_2$, 25.3 µM ZnO, 0.08 µM Na_4MoO_4, 111 mM glucose and 60 mg/l for each of the 19 natural amino acids and 40 mg/l for tyrosine (Sigma-Aldrich, St Louis, USA). For temperature switch experiments two cultures (A and B) were grown aerobically at 37 °C in M9 minimal medium as described above, in order to have a biological replicate. After overnight culture, cells were inoculated at OD_{600} 0.01 and allowed to grow into exponential phase before being harvested at OD_{600} 1.0. These cells were washed at room temperature with complete M9 minimal medium (but lacking methionine) to prevent osmotic shock during washing and then transferred to M9 minimal medium in which the methionine was replaced by 400 mg/l azhal. The cultures were split and then transferred to water-bath shakers set to either 37 or 44 °C. The four cultures were allowed to resume growth aerobically for 15 minutes before being harvested.

Sample preparation— Samples were essentially prepared as described before (141), in short: azhal-labelled cells were harvested by centrifugation, pellets were resuspended in lysis buffer and lysed by sonication after which cellular debris was removed by centrifugation. Samples were dialyzed overnight and their protein content was determined. Samples were then subjected to overnight digestion with trypsin.

iTRAQ labelling— Samples (125 µg protein per sample) were lyophilized after digestion and redissolved in 40 µl of 125 mM tri-ethyl-ammonium bicarbonate pH 8.5, and labelled with iTRAQ (244) according to the manufacturer's protocol (Applied Biosystems, Toronto, Canada), with the exception that two vials of iTRAQ reagent were used per sample to ensure complete labelling. Samples were incubated for two hours at room temperature after which the reaction was quenched by adding 300 µl of 0.1% formic acid. The digests from the cultures A and B grown at 37 ºC during pulse-labelling were labelled with iTRAQ 114 and 116 respectively, while digests from the cultures A and B grown at 44 ºC were labelled with iTRAQ 115 and 117. The four labelled samples were mixed in a 1:1:1:1 (w/w) ratio, resulting in 500 µg iTRAQ labelled digest. To remove the excess of iTRAQ reagent the sample was diluted three times to a final volume of 6 ml 20% acetonitrile in 0.1% formic acid and loaded on an ICAT cation exchange cartridge (Applied Biosystems, Toronto, Canada). The cartridge was washed with 500 µl 20% acetonitrile in 0.1% formic acid, before the digest was eluted with 2 M ammonium formate buffer pH 6.8 containing 20% acetonitrile and lyophilized. Samples were redissolved in 50 mM Hepes pH 8.0 and reduced and alkylated as described above, before ~200 µg was loaded for the primary run of diagonal chromatography as described below.

COFRADIC and mass spectrometric analysis—Azhal-containing peptides were enriched by COFRADIC (213), using TCEP to selectively modify target peptides between the primary and secondary chromatographic runs. TCEP induces a set of competing reactions in azhal-containing peptides present in primary fractions, i.e., conversion of the azido-group to an amine or hydroxyl group and cleavage of the peptide bond at the C-terminal side of azhal residues (141, 143, 223). The subsequent enrichment is based on a difference in retention-time during the secondary chromatographic runs between TCEP-induced reaction products and the bulk of unmodified peptides that are present in the particular primary fraction subjected to TCEP treatment (141). Three fractions 16 minutes apart in the primary run are pooled and reinjected after TCEP treatment. Fractions of secondary runs enriched in TCEP-induced reaction products from azhal-containing peptides, were analyzed by LC tandem-MS. These 'Off-diagonal' pooled

fractions were redissolved in 10 µl 0.1% TFA with the addition of 150 pmol human [Glu1]-Fibrinopeptide B (Sigma-Aldrich, St Louis, USA) for internal calibration, 10 µl sample was separated and analyzed as described in detail previously (141). In addition pooled fractions containing non-shifted material were also collected, resuspended in 400 µl of 0.1% TFA with the addition of 150 pmol of human [Glu1]-Fibrinopeptide B (Sigma-Aldrich) for internal calibration. Of these samples 5 µl was injected and analysed as described for the fractions enriched with labelled material. Assessment of the relative quantity of each protein by analysis of tandem-MS spectra is based on the signal intensities of reporter ions derived from the iTRAQ-moieties of the respective peptides.

Identification and Quantitation— The tandem-MS runs were first internally recalibrated on the fragmentation spectrum of [Glu1]-Fibrinopeptide B, before being exported by the mascot.dll as described before (141). Generated peak lists were submitted to MASCOT to identify newly synthesized proteins using the following parameters: Cleavage after lysine or arginine unless followed by proline plus cleavage after methionine, allowing up to 2 missed cleavages, fixed carbamidomethyl cysteine, iTRAQ (K) modifications. Variable modifications used were iTRAQ (N-terminal) modification and modifications induced by reaction of TCEP with azhal-containing peptides as described before (141). Peptide mass tolerance was set at 0.1 Da and MS/MS tolerance was set at 0.05 Da. The significance threshold was set to 0.01 resulting in a threshold score of 34. Mudpit scoring and 'require bold red' were applied with an ion-score cut-off of 35, in order to have all peptide matches identified at a p-value of <0.01. Mascot performed fragment ion searches with the above settings in a local database of the *E. coli* K12 proteome (4506 proteins; 1426768 residues, release 14.4 04/11/08, Uniprot consortium, http://beta.uniprot.org/). In addition peak lists from non-shifted fractions was searched with the following parameters: cleavage after lysine or arginine unless followed by proline, allowing up to one missed cleavage, fixed carbamidomethyl cysteine modification, iTRAQ (K) modification. Variable modification used was iTRAQ (N-terminal) modification. Peptide mass tolerance was set at 0.1 Da, and MS/MS tolerance was set at 0.1 Da. The significance threshold was set to 0.01 resulting in a threshold score of 28. Multidimensional protein identification technology scoring and "require bold red" were applied with an ion score cut-off of 35 to have all peptide matches identified at a *p* value of <0.01. MASCOT, False positive rates were estimated using a decoy database as described before (141) and were found to be less than three percent. TCEP-induced reaction products were selected manually and selected queries were used to recalculate protein coverage and protein score based on azhal-labelled peptides only. The resulting MASCOT data-file of this search was imported into Quant (245) (http://www.protein-ms.de), for quantitation using the iTRAQ reporter ions using only labelled peptides unique to each protein. For identification purposes both searches were also exported as csv-files. Quant settings were as follows: all four iTRAQ reporters on, report peak areas on, reporter tolerance set at 0.1 Da, intensity range turned of, peak dimensions at 0.025 Da, absolute intensity error set at zero, experimental error set at 0%, use of unique peptides on, p-value cut-off set at 0.01 and macro language set at English with the macro parameter separator set to comma. The correction factors were put in for iTRAQ kit no. 080591.

Quant output is a tab-delimited text file containing both reporter ion ratios per peptide as well as mean-protein ratios derived from these. To assess the combined effect of technical and biological variance the average and standard deviations of all protein ratios of 116/114 and 117/115 for the data-set were calculated. These ratios represent replicate B/replicate A at 37 ºC and replicate B/replicate A at 44 ºC and should theoretically be one. The 116/114 ratio yielded an average of 1.01 with standard deviation (s.d.) 0.19 and the 117/115 ratio an average of 1.03 (s.d. 0.20), for the azhal-labelled proteins, while for the unlabeled proteins the 116/114 ratio yielded an average of 1.07 (s.d. 0.17) and the 117/115 ratio an average of 1.11 (s.d. 0.18), which shows that no large systemic error was made during mixing of samples. Subsequently tandem-MS spectra were inspected manually. Peptides which did not have signals for all four reporter ions or peptides which showed inconsistent 115/114 and 117/116 ratios (biological plus technical replicates) were discarded. Box-plots per protein were checked to identify outliers in the peptide ratios per protein as described in (245). In addition for quantitation of relative total protein level changes induced by temperature, methionine containing peptides were removed from the non azhal-containing peptide set as well.

Due to the incorporation of replicates into one quantitation-experiment, each peptide can yield four relevant reporter ratios. First of all: 115/114 and 117/116 for each pre-labelling culture split into two different growth temperatures during labelling. Furthermore 117/114, 115/116 for one growth temperature of one pre-labelling culture compared to the other growth temperature of the other pre-labelling culture during labelling. For each protein

the mean expression ratio was determined by calculating the mean of the peptide ratios 115/114, 117/114, 115/116 and 117/116 per protein (mean ratio of expression between the biological replicates). The accuracy of the ratio per protein is expressed by calculating the standard deviation of the peptide ratios for all the peptides measured per protein (maximum standard deviation between the biological replicates). To ascertain if up- or down-regulation was significant, a double sided Welch's T-test for each protein was performed, using the Welch-Satterthwaite equation (246) to determine the degrees of freedom, to see whether the mean protein ratio differed significantly from the mean ratio obtained for unregulated proteins ($\mu = 1.02$, $\sigma = 0.16$, n=343 for newly synthesized proteins and $\mu = 1.08$, $\sigma = 0.17$, n=291 for protein levels; from ratio of 116/114 and 117/115 reporter ions). Proteins which changed more than 1.5 fold and for which the p-value was adjusted for a false discovery rate of less than <0.05 due to multiple testing (247) were considered to have a significantly altered expression level.

5

Proteome-wide alterations in *E. coli* translation rates upon anaerobiosis

Summary

Enzyme reprofiling in bacteria during adaptation from one environmental condition to another may be regulated both by transcription and translation. However, little is known about the contribution of translational regulation. Recently we have developed a pulse-labelling method using the methionine analogue azhal to determine the relative amounts of proteins synthesized by *E. coli* in a brief time frame upon a change in environmental conditions. Here we show that measuring changes in total protein levels on the same time-scale as new protein synthesis allows identification of stable and labile proteins. We demonstrate that altered levels of most newly synthesized proteins are the result of a change in translation rate rather than degradation-rate after a switch from aerobiosis to anaerobiosis. The majority of proteins with increased synthesis-rates upon an anaerobic switch are involved in glycolysis and pathways aimed at preventing glycolysis grinding to a halt by a cellular redox-imbalance. Our method can also be used to compare relative translation rates with relative mRNA levels. Discrepancies between these parameters may reveal genes whose expression is regulated by translation. This may help unravelling molecular mechanism underlying regulation of translation, e.g. mediated by small regulatory RNAs.

Introduction

An integrated view of the molecular events underlying adaptation of bacteria to major environmental changes requires insight into both transcriptional and post-transcriptional regulation of gene expression. With the availability of annotated genome databases, much has been learned about global changes induced in mRNA levels and the transcription factors involved, as well as about changes in steady state protein levels upon a switch in environmental conditions. However, little is known concerning genome-wide changes in protein synthesis and degradation-rates and about the contributions of transcription and translation to the regulation of gene expression when bacteria adapt to major changes in the environment. An important reason for this is the fact that it is much easier, using genome-wide microarray analysis, to find candidate genes with expression levels being regulated via RNA transcription or degradation, than to identify candidate genes with expression levels regulated at the level of translation.

Renewed interest in gene expression regulation at the post-transcriptional level in prokaryotes has been sparked recently by the discovery in 2001-2002 of the existence of large numbers of small regulatory RNAs (sRNAs) in *E. coli* and other bacteria (248-251). These sRNAs may regulate translation of numerous mRNAs (61), often mediated by the RNA chaperone Hfq (252). Although the function of many sRNAs is not yet known, several recently reviewed findings (62) strongly suggest that post-transcriptional regulation by sRNAs is widespread and that the number of mRNAs regulated by sRNAs amply exceeds the number of sRNAs themselves. However, not much is known about how environmental signals are transduced to sRNA mediated regulation of translation (57). Identifying potential target genes is an important step in unravelling underlying molecular mechanisms of translational regulation.

What is needed is a proteomic method to determine alterations in translation rates during adaptation to environmental changes. By comparison of changes in translation rates with changes in mRNA levels genes may be identified that are regulated at the translational level. Recently we have developed a pulse-labelling technique using the methionine analogue azhal. This enables assessment of the relative amounts of proteins synthesized in a brief period during adaptation to a major environmental change, on a proteomic scale (141). We previously applied this approach to examine early changes in newly synthesized proteins upon a sudden rise in growth temperature. However, the changed levels of newly synthesized proteins after a pulse of several minutes could be the result of a change in either synthesis or degradation-rate or a combination of both. A prerequisite to identify proteins with altered translation rates is that the protein half-life far exceeds the labelling time used. The vast majority of abundant proteins in growing *E. coli* cells is relatively stable (109) and has a half-life of at least a few hours (110, 111), while a small pool is rapidly degraded (90). However, no proteome-wide information on protein turnover for individual proteins in *E. coli* is available. Here we show that our analytical strategy to determine relative amounts of newly synthesized proteins by pulse-labelling with azhal can be easily extended to identify stable and labile proteins. Since

degradation of stable newly synthesized proteins during a 10 minutes pulse is negligible, increased or decreased newly synthesized amounts of stable proteins is predominantly the result of a change in translation rate rather than in degradation-rate.

In this study we use the extended azhal pulse-labelling approach and quantitative mass spectrometry to determine, on a proteomic scale, changes in both newly synthesized proteins as well as total protein levels in exponentially growing *E. coli* cells during their initial adaptation to a sudden drop in oxygen levels. By comparison of changes in newly formed proteins to changes in total protein levels in the initial 10 minutes following the anaerobic switch, stable and labile proteins are identified. For the large group of stable proteins the relative changes in levels of newly synthesized proteins directly reflect the average relative translation rates of these proteins over the pulse-labelling time used. We obtained a dataset of relative translation rates that is consistently related to the required metabolic adaptation, underscoring the reliability of our method.

Results

Growth, labelling and quantitation of newly synthesized proteins upon an anaerobic switch— *E. coli* is a facultative anaerobic prokaryote and its ability to switch between aerobic and anaerobic environments greatly expands the range of niches it can thrive in. To enable growth in both environments, a set of metabolic routes for each has to be maintained. Changes that occur in the central carbon metabolism at the onset of anaerobiosis are related to the need for alternative ways to maintain a proper intracellular redox-balance, since molecular oxygen, the terminal electron acceptor of the aerobic electron transport chain, is no longer available to remove the excess reducing equivalents formed in catabolism. This can be achieved by anaerobic respiration, using one or more alternative terminal electron acceptors (253). If no alternative electron acceptor is available, growth on glucose requires removal of reducing equivalents by mixed acid fermentation (254). When grown in aerobic batch cultures, with glucose as energy source and in the absence of alternative electron acceptors, *E. coli* stops growing upon a switch to anaerobiosis for about 20 minutes and then resumes growth at a slower pace (124). Under these conditions the flux through the phosphotransferase system (PTS) for glucose transport annex phosphorylation and through glycolysis is dramatically increased (255). To study immediate changes in protein synthesis upon an anaerobic switch, methionine-auxotrophic cells grown aerobically in an amino acid supplemented glucose-containing medium were harvested, washed with the same medium without methionine and transferred to either an aerobic or anaerobic environment and pulse-labelled for 10 minutes with azhal. During the pulse of azhal, aerobically cultured cells continue growth and increase their cellular mass by ~6% in 10 minutes, judged from an increase in optical density at 600 nm, while the anaerobic cells stop growing, in agreement with earlier observations (124)

Proteins extracted from azhal-labelled cultures were digested and peptides were labelled with iTRAQ for quantitation. Subsequently, azhal-containing peptides were enriched by COFRADIC, and identified and quantified by LC tandem-MS as outlined in *experimental procedures* and previously described (141). A total of 414 azhal-containing peptides were identified, corresponding to 211 different proteins synthesized during the pulse-labelling period. From the 211 newly synthesized proteins, 164 could be quantified using the iTRAQ reporter ions, according to the criteria referred to in *experimental procedures*. Upon a switch to an anaerobic environment the relative abundance of 69 newly synthesized proteins significantly ($p<0.05$) increased or decreased more than twofold. In addition 31 proteins changed significantly ($p<0.05$) by a factor of 1.5-2. The relative abundance of the remaining 64 newly synthesized proteins changed less than 1.5 fold or did not change significantly at all during the 10 minutes after the anoxic switch (Figure 1).

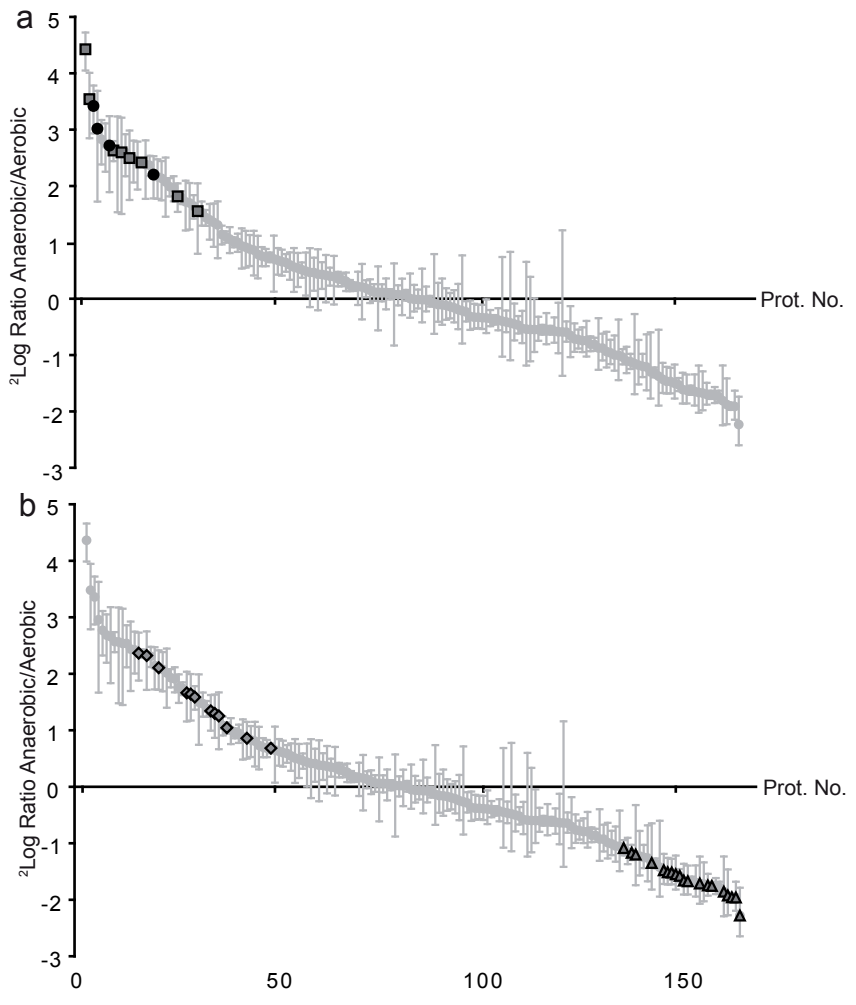


Figure 1. **Relative levels of proteins newly synthesized during pulse-labelling with azhal.** Light grey dots, ^{2}log ratio of newly formed proteins ordered from most increased to most decreased under anaerobic conditions. Error bars indicate standard deviations. Proteins belonging to different pathways are listed in tables I and II. Panel (**a**) dark grey squares, proteins involved in anaerobic respiration; black dots, mixed acid fermentation enzymes. Panel (**b**): grey diamonds, glycolytic enzymes and PTS system proteins; dark grey triangles, ribosomal proteins.

Pathways affected in the rate of protein synthesis by a sudden change to an anaerobic environment— The change to an anaerobic environment necessitates either activation of the anaerobic respiration machinery or of fermentation pathways. Five alternative terminal reductases can be expressed in *E. coli* (253). We detected newly synthesized subunits from four of these, nitrite-reductase (NirB), DMSO-reductase (DmsA), nitrate reductase (NarG) and fumarate-reductase (FrdA), all being increased at least five-fold. It is noteworthy that the synthesis of NirB, DmsA and NarG rapidly increases while their respective substrates are not present in the environment. The relative new formation under anaerobic conditions of two proteins (MoaB and MoaE) involved in the biogenesis of a molybdenum-cofactor was also increased. This cofactor is part of different molybdenum-containing enzymes, amongst which are the anaerobic respiratory enzymes TMAO-reductase, nitrate-reductase and DMSO-reductase (256). Furthermore we find increased amounts of newly synthesized cytochrome-d oxidase (CydA) and a putative quinone oxidoreductase (YhdH) (257). The cytochrome-d oxidase is known to be expressed under conditions with low oxygen tensions (258), while increased synthesis of the putative quinone oxidoreductase suggests a function in anaerobic respiration.

Apart from anaerobic respiration, *E. coli* can maintain its redox balance under anaerobic conditions via oxidation of reducing equivalents by mixed acid fermentation (Figure 1a, black dots and Table I). Pyruvate formate lyase (PFL), together with its auxiliary glycyl radical cofactor (GrcA) (259), plays a central role in this respect as it supplants pyruvate dehydrogenase (PDH) in converting pyruvate. PFL catalyzes the formation of formate and acetyl-CoA from pyruvate. Thus, PFL generates both a fermentation end-product and a precursor which can be converted to acetate in a reaction coupled to formation of ATP, or undergo further fermentation to ethanol. Not surprisingly, it was found that the amounts of newly synthesized PflB and GrcA are highly increased upon the anaerobic switch. Concordantly, two major enzymes that regenerate NAD^+ from NADH in fermentation are also strongly induced upon the anaerobic switch, namely alcohol-dehydrogenase (AdhE) that catalyzes the conversion of acetyl-coA to ethanol and D-lactate-dehydrogenase (LdhA) that reduces pyruvate to lactate. Not all proteins involved in fermentation show dramatic upregulation in synthesis during the first ten minutes upon anaerobic switching. AckA, involved in the conversion of acetyl-CoA to acetate, is only modestly up-regulated during the initial response to an anaerobic environment. So, the relative levels of four newly synthesized enzymes (PflB, GrcA, AdhE and LdhA) out of the six enzymes involved in the conversion of pyruvate to formate, acetate, ethanol and lactate were increased at least fourfold, while one (AckA) was increased slightly and the other one (Pta) was not observed. It is also noteworthy that levels of the newly synthesized PDH complex subunits AceE, AceF and LpdA hardly change upon anaerobiosis, even though PDH activity is strongly inhibited under anaerobic conditions (255).

Another major pathway that is affected by a switch to anaerobic growth in the absence of alternative electron acceptors is glycolysis. It is the pathway that is responsible

TABLE I
Identity of newly synthesized proteins most elevated 10 minutes after a switch to anaerobiosis in E. coli.

protein name	gene	ratio†	S.D.	ratio‡	S.D.	pathway
nitrite reductase [NAD(P)H] large subunit	*nirB*	20.6	4.7	n.d.		*
anaerobic dimethyl sulfoxide reductase	*dmsA*	11.2	4.2	n.d.		*
aldehyde-alcohol dehydrogenase	*adhE*	10.3	2.9	2.6	0.68	#
pyruvate formate-lyase	*pflB*	7.8	4.6	2.3	0.73	#
universal stress protein D	*yiiT*	6.8	1.8	n.d.		
universal stress protein A	*uspA*	6.4	1.9	1.5	0.17	
autonomous glycyl radical cofactor	*grcA*	6.3	2.7	5.4	1.6	#
glycerol dehydrogenase	*gldA*	5.9	0.72	n.d.		
universal stress protein G	*ybdQ*	5.9	3.1	n.d.		
respiratory nitrate reductase 1	*narG*	5.8	3.1	n.d.		*
ketol-acid reducto isomerase	*ilvC*	5.8	1.5	n.d.		
fumarate reductase flavoprotein subunit	*frdA*	5.4	2.2	n.d.		*
ribosome-associated inhibitor A	*raiA*	5.4	1.3	n.d.		
6-phosphofructokinase isozyme 1	*pfkA*	5.2	1.5	1.8	0.27	•
putative quinone oxidoreductase yhdH	*yhdH*	5.1	0.27	n.d.		*
fructose-specific phosphotransferase IIA	*fruB*	5.0	1.7	n.d.		•
7-alpha-hydroxysteroid dehydrogenase	*hdhA*	5.0	0.61	0.95	0.07	
D-lactate dehydrogenase	*ldhA*	4.4	1.1	n.d.	.	#
phosphoglyceromutase	*gpmI*	4.3	1.0	1.4	0.10	•
bacterioferritin	*bfr*	4.2	1.0	n.d.		
probable sigma (54) modulation protein	*yhbH*	4.0	1.4	n.d.		
malate synthase G	*glcB*	3.8	0.42	n.d.		
molybdenum cofactor biosynthesis protein B	*moaB*	3.8	0.57	n.d.		
cytochrome d ubiquinol oxidase subunit 1	*cydA*	3.4	0.59	1.7	0.24	*
small heat shock protein ibpB	*ibpB*	3.3	0.26	n.d.		
phosphoglycerate kinase	*pgk*	3.2	0.94	1.3	0.21	•
pyruvate kinase II	*pykA*	3.1	0.87	1.2	0.16	•
glucose-6-phosphate isomerase	*pgi*	3.0	0.43	1.2	0.38	•
molybdopterin-converting factor subunit 2	*moaE*	2.8	0.41	n.d.		
UPF0265 protein yeeX	*yeeX*	2.6	0.25	1.4	0.30	
phosphoenolpyruvate-proteinphosphotransferase	*ptsI*	2.5	0.55	1.1	0.13	•
glyceraldehyde-3-phosphate dehydrogenase A	*gapA*	2.5	0.64	1.2	0.16	•
enolase	*eno*	2.4	0.80	1.1	0.21	•
6-phosphogluconolactonase	*pgl*	2.1	0.24	n.d.		
fructose-bisphosphate aldolase class 2	*fbaA*	2.1	0.14	1.1	0.13	•
threonyl-tRNA synthetase	*thrS*	2.0	0.32	1.1	0.18	

†, ratio of the amounts of proteins newly synthesized for 10 minutes upon the onset of anaerobiosis and under aerobic conditions determined by quantitation of azhal-containing peptides; ‡, ratio of the total protein levels (pre-existing plus newly synthesized) 10 minutes after the onset of anaerobiosis and under aerobic conditions determined by quantitation of peptides lacking both azhal and methionine; *, proteins involved in anaerobic respiration; #, proteins involved in fermentation; •, glycolytic enzymes and PTS sugar transport proteins; n.d., not determined.

for the generation of most energy from glucose under these conditions. Figure 1b (grey diamonds) shows that the relative amounts of the 9 detected glycolytic enzymes synthesized during the first ten minutes after the anaerobic switch increase between 1.6 and 5.2 fold (Table I). This represents ~ 90% of the entire pathway, the only glycolytic enzyme not observed being triose isomerase. Our results corroborate and greatly extend previous observations regarding increases in levels of newly synthesized glycolytic and mixed acid fermentation enzymes (124).

In addition, we observe an increase in the amounts of newly synthesized glucose-specific (PtgA) and fructose-specific (PtfaH) components as well as the phosphoenolpyruvate-protein phosphotransferase (PtsI) part of the PTS which is involved in sugar import and its subsequent phosphorylation prior to entering glycolysis (Figure 1b, grey diamonds and Table I). The PTS protein HPr is lacking in our dataset, probably because tryptic digestion will only yield one methione-containing peptide. This peptide of 6 amino acids, including the protein C-terminus, can easily escape detection. We did also not detect one other PTS protein, PtsG, a membrane protein responsible for transport and concomitant phosphorylation of glucose. Despite the absence of some proteins in our dataset, these results indicate that the relative amount of most if not all newly synthesized members of the entire PTS and glycolysis have increased considerably 10 minutes after the onset of anaerobiosis.

The synthesis of three uniform stress proteins (260) i.e. uspD, uspA and uspG was found to increase about six-fold upon anaerobiosis. Two ribosome-associated factors, RaiA and YhbH, were also produced in much higher amounts under anaerobiosis. These two proteins are also found in cells in stationary phase (261) and their upregulation may be related to the transient growth arrest during adaptation to the anaerobic state.

Because transient growth arrest occurs in the anaerobic environment, the overall protein synthesis-rate is expected to become lower under these conditions. Indeed, the levels of all 20 detected newly synthesized ribosomal proteins (out of total of 56), the expression of which is strongly related to growth rate (262, 263), were significantly lower in the anaerobic cells (dark grey triangles in Figure 1b and Table II). Because of their abundance, the synthesis-rates of ribosomal proteins contribute considerably to the overall protein synthesis-rate. Overall, both the proteins found to be up-regulated as well as those found to be down-regulated immediately following the anoxic switch are related to the imposed change in environment. This functional consistency underscores the reliability of our dataset.

TABLE II

Identity of newly synthesized proteins most decreased 10 minutes after a switch to anaerobiosis in E. coli.

protein name	gene	ratio†	S.D.	ratio‡	S.D.	pathway
50S ribosomal protein L11	*rplK*	0.21	0.06	0.95	0.27	#
50S ribosomal protein L4	*rplD*	0.26	0.05	0.86	0.13	#
30S ribosomal protein S8	*rpsH*	0.26	0.01	0.91	0.07	#
30S ribosomal protein S10	*rpsJ*	0.26	0.07	0.92	0.17	#
50S ribosomal protein L2	*rplB*	0.28	0.10	0.86	0.13	#
cysteine synthase A	*cysK*	0.29	0.02	0.51	0.08	
di-aminopimelate decarboxylase	*lysA*	0.29	0.03	n.d.		
30S ribosomal protein S7	*rpsG*	0.30	0.01	0.98	0.17	#
30S ribosomal protein S20	*rpsT*	0.30	0.04	0.83	0.17	#
peptide deformylase	*def*	0.30	0.07	n.d.		
50S ribosomal protein L1	*rplA*	0.31	0.09	0.89	0.15	#
peptidoglycan-associated lipoprotein	*pal*	0.31	0.06	0.74	0.13	
GTP cyclohydrolase 1	*folE*	0.31	0.02	n.d.		
50S ribosomal protein L6	*rplF*	0.31	0.06	0.94	0.08	#
50S ribosomal protein L13	*rplM*	0.32	0.06	0.95	0.11	#
50S ribosomal protein L9	*rplI*	0.34	0.03	0.93	0.16	#
30S ribosomal protein S9	*rpsI*	0.34	0.07	0.92	0.18	#
30S ribosomal protein S1	*rpsA*	0.35	0.05	1.02	0.20	#
50S ribosomal protein L3	*rplC*	0.35	0.06	0.85	0.15	#
50S ribosomal protein L17	*rplQ*	0.36	0.07	0.83	0.11	#
alkyl hydroperoxide reductase subunit F	*ahpF*	0.37	0.16	0.49	0.14	
quinoprotein glucose dehydrogenase	*gcd*	0.39	0.02	n.d.		
50S ribosomal protein L5	*rplE*	0.39	0.15	0.90	0.13	#
biopolymer transport protein exbB	*exbB*	0.42	0.14	n.d.		
beta-hydroxydecanoyl thioester dehydrase	*fabA*	0.42	0.06	0.93	0.10	
polyribonucleotide nucleotidyltransferase	*pnp*	0.43	0.11	0.75	0.17	
30S ribosomal protein S5	*rpsE*	0.43	0.20	0.90	0.12	#
50S ribosomal protein L21	*rplU*	0.44	0.04	0.90	0.16	#
DNA-directed RNA polymerase subunit beta	*rpoB*	0.45	0.07	0.96	0.15	
30S ribosomal protein S2	*rpsB*	0.47	0.04	0.93	0.13	#
magnesium-transporting ATPase, P-type 1	*mgtA*	0.48	0.17	n.d.		
cold-shock DEAD box protein A	*deaD*	0.48	0.10	n.d.		
uridylate kinase	*pyrH*	0.50	0.07	n.d.		

†, ratio of the amounts of proteins newly synthesized for 10 minutes upon the onset of anaerobiosis and under aerobic conditions determined by quantitation of azhal-containing peptides; ‡, ratio of the total protein levels (pre-existing plus newly synthesized) 10 minutes after the onset of anaerobiosis and under aerobic conditions determined by quantitation of peptides lacking both azhal and methionine; # ribosomal proteins, n.d. not determined.

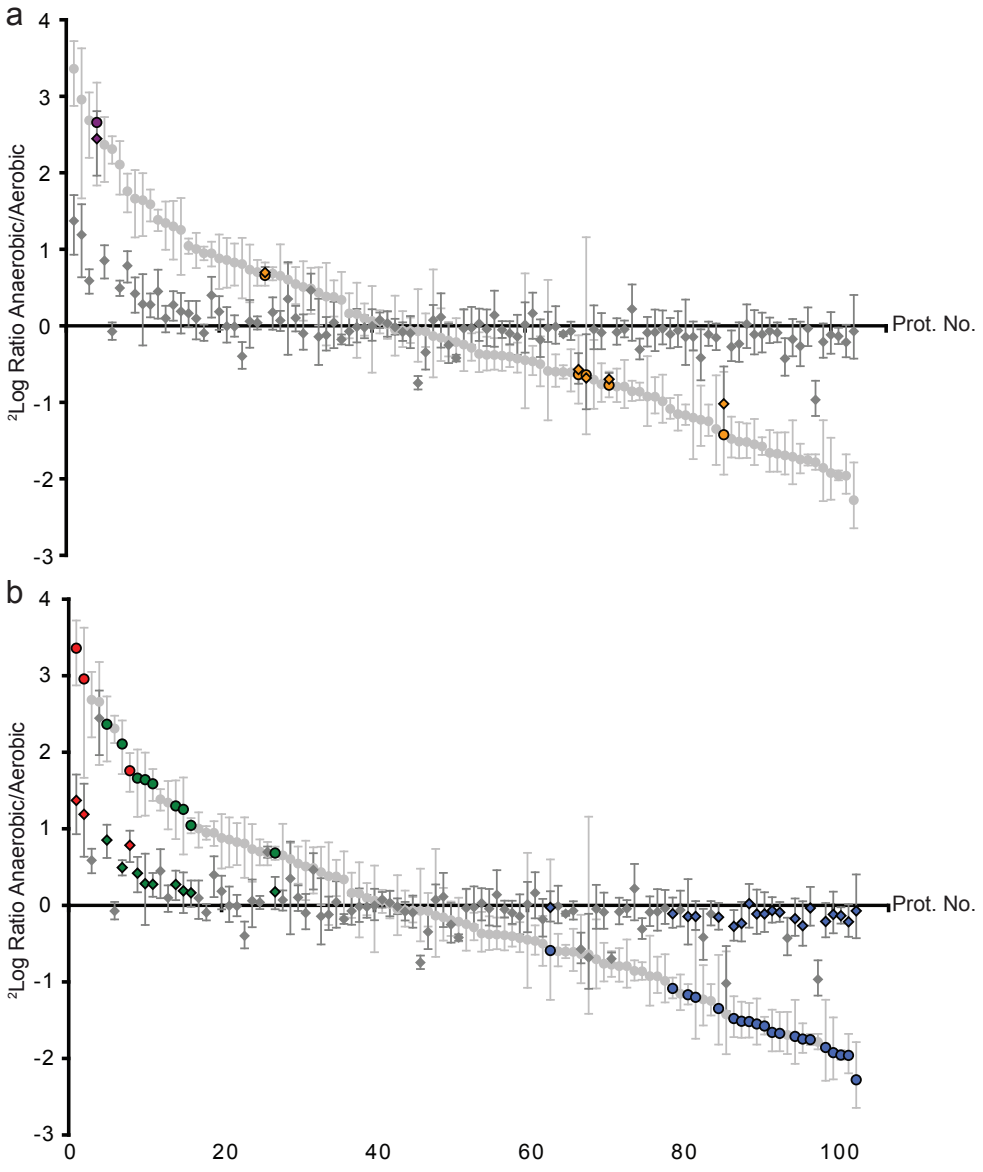


FIGURE 2. **Changes in protein synthesis compared to changes in protein levels upon onset of anaerobiosis.** Light grey or coloured dots, relative change of newly synthesized proteins; dark grey or coloured diamonds at the same position on the X-axis as a dot are relative total protein levels of the same protein 10 minutes after the switch to anaerobiosis. Proteins are ordered from most induced to most repressed synthesis. In panel (**a**) proteins having a *high* turnover rate are marked, namely GrcA (purple dot and diamond) and ClpA, KatG, AhpC, MetN and AhpF (orange dots and diamonds). In panel (**b**) coloured symbols are used to represent proteins with a *low* turnover rate in the same data set. Red dots and diamonds, proteins involved in anaerobic respiration and mixed acid fermentation; green dots and diamonds, glycolysis; blue dots and diamonds, ribosomal proteins.

Changes in newly synthesized proteins are predominantly the result of an altered translation rate rather than an altered degradation-rate— Changes in levels of newly synthesized proteins after a pulse of 10 minutes could be the result of a change in synthesis-rate, degradation-rate or both. The vast majority of proteins in growing *E. coli* cells is relatively stable (109) with half-lives that exceed 2h (110, 111). Proteins of the small pool that is rapidly degraded often have regulatory functions (90). However, no proteome-wide information on protein turnover for each individual protein is available.

To identify proteins of which the change in expression is predominantly the result of a change in translation rate rather than a change in degradation-rate we also analyzed the 'un-shifted' fractions obtained during COFRADIC by LC tandem-MS. Peptides in these fractions that do not contain azhal or methionine can be used to quantify changes in total protein levels during the pulse, since they originate from both pre-existing and newly synthesized proteins. In addition, shifted reversed phase HPLC fractions, enriched for azhal-containing peptides, also contain peptides without azhal after secondary runs of COFRADIC. Their presence is due to tailing of main chromatographic peaks that contain the bulk of unlabeled non-shifting peptides. The combined database searches from data acquired from shifted and non-shifted pooled fractions yielded a total of 1451 peptides not containing azhal or methionine, identifying 344 proteins. Of these a total of 305 proteins could be quantified using the reporter ions from iTRAQ. Ten minutes after a switch to an anaerobic environment the total relative amounts of 6 proteins had changed significantly ($p<0.05$) by a factor of two or more, while 12 proteins changed 1.5-2 fold ($p<0.05$) and 287 proteins changed less than 1.5 fold or had not changed significantly. Increases in total protein levels within ten minutes after the switch were detected for enzymes involved in fermentation, anaerobic respiration, glycolysis, and PTS (Supplemental Figure 1).

Our datasets of azhal/methionine-lacking and azhal-containing peptides contained 103 proteins of which both the relative total levels during the pulse of 10 minutes and the relative levels of newly synthesized species had been determined. The two datasets are combined in tables I and II and in Figure 2. In Figure 2 the data are ordered according to the relative levels of newly synthesized proteins (dots), while corresponding total levels are represented by diamonds at the same position along the X-axis. As one would expect, the relative total levels (diamonds) of the great majority of proteins change much less than the corresponding relative levels of newly synthesized species (dots). However, we also identified six proteins, GrcA, ClpA, KatG, AhpC, MetN and AhpF, of which the relative total levels changed significantly upon anaerobiosis, almost as much as the relative levels of the newly synthesized species (Figure 2a). Apparently, practically all pre-existing protein molecules of this group have been replaced by newly synthesized polypeptides during the pulse. So, the half-life of these proteins lies well within the pulse-labelling time used. The high turnover of GrcA (115) and ClpA (264, 265) have been found before, while catalase (KatG) has been identified as a putative substrate for the protease ClpB (100). Interestingly AhpC, subunit of a peroxidase, is also identified as a labile protein in a dataset of relative levels of

newly synthesized and total protein after a change in growth temperature (*Chapter 4*) under conditions as described before (141), which confirms the observations during the switch to anaerobiosis. Changes in KatG and AhpF are small under these conditions of heat shock. However, changes in protein synthesis of AhpF still seem to be mirrored by changes in total protein levels on the same time-scale. Among the other proteins identified to be labile proteins following heat shock, are the heat shock sigma factor σ^{32} (Rp32) and IbpB, a member of the family of small heat shock proteins, both known to have a high turnover rate (*Chapter 4*), further confirming that the observations relate to turnover rate of the proteins measured. For these labile proteins the change in the level of newly synthesized species could be caused by a change in synthesis-rate, degradation-rate or a combination of the two.

In contrast to the proteins for which the relative total levels change to almost the same degree as the relative newly synthesized levels, a significant difference between relative total levels and relative levels of newly synthesized material after a pulse of 10 minutes was noted for the majority of proteins. This is shown in Figure 2b by red symbols for mixed acid fermentation and anaerobic respiration enzymes, by green symbols for the glycolytic enzymes and proteins of the PTS system, and by blue symbols for ribosomal proteins. For the ribosomal proteins the same trend was again observed during a change in growth temperature (*Chapter 4*). The difference between levels of total and newly synthesized protein can be explained by assuming that these proteins have a half-life that (far) exceeds the pulse-labelling time used. Consequently, a large change in synthesis will not affect the total protein levels to the same extent in the short time frame used for pulse-labelling; only a continued altered rate of formation will eventually change the protein levels to the same extent. For these stable proteins the effect of degradation-rate is negligible in the pulse-labelling time used and the relative changes in newly synthesized species directly reflect the relative translation rates of these proteins during the pulse of cells grown in aerobic or anaerobic conditions. An example is the 13.5 fold increase of average synthesis-rate found for PflB during 10 minutes after the switch to anaerobiosis by Smith *et al.* (124) using radiolabelling, which is in good agreement with the 7.8 fold upregulation of the synthesis-rate found with azhal pulse-labelling.

A model for changes in newly synthesized protein levels related to changes in total protein levels during the pulse is presented in the supplemental data. It is clear from the example using glycolytic enzymes (Supplemental Table I) that the measured changes in total levels and in protein synthesis-rate is best approximated by assuming a low turnover (half-life ~2h) for these proteins. This indicates that the increased level of newly synthesized glycolytic enzymes during the pulse is predominantly the result of an increased rate of protein synthesis rather than a decrease in the rate of degradation. In general, the relative amounts of most newly synthesized proteins reflect the average relative translation rates during the pulse.

Discussion

This study demonstrates how the proteome-wide azhal pulse-labelling technique combined with COFRADIC enrichment can be used to obtain reliable quantitative data regarding newly synthesized proteins during initial phases of environmental transitions. Although the approach thus far has been based on the subset of methionine containing peptides to identify and quantify newly synthesized proteins, no bias towards proteins with higher methionine content was found compared to the standard approach when all tryptic peptides are employed. Extension of the approach to also determine changes in total protein levels on the same time-scale as new protein synthesis, allows identification of stable and labile proteins. Turnover was determined to be low for most proteins that exhibited a sharp change in new protein formation upon anaerobiosis. This is consistent with earlier notions on protein turnover in *E. coli* (90, 109-111) and correlates well with previous reports on protein turnover in other organisms (116, 119). On the whole this indicates that the change in levels of newly synthesized, stable proteins is caused by a change in protein translation rate, as degradation does not contribute significantly to changes in levels of new protein formation, on the time-scale employed.

The early response of *E. coli* to an anaerobic environment was found to consist of an immediate and strong increase in synthesis-rate of proteins involved in anaerobic respiration and fermentation. Furthermore the synthesis-rates of glycolytic enzymes and PTS proteins increased as well. This strongly suggests that the increased flux through glycolysis (255) is not exclusively the result of metabolic regulation but is also regulated at the protein level. Most of the proteins found to be down-regulated in synthesis were ribosomal proteins. Their rate of synthesis is strongly related to growth rate (262, 263). It has been well documented that under conditions of energy deficiency synthesis of rRNA is decreased (266), leading to translational feedback inhibition of ribosomal protein synthesis (53). The early changes in synthesis-rate upon the anaerobic switch for some of the proteins measured seem to be a good early indication of the steady state levels measured under anaerobic growth, as can be derived from the study of Smith *et al.* (124) for enolase, pyruvate kinase I, pyruvate-formate lyase B and glyceraldyde 3-phosphate dehydrogenase. Consequently, measuring changes in the rate of synthesis seems to be a more sensitive indicator of regulation on a short time-scale than determination of changes in protein levels, especially for stable abundant proteins, as is shown by the small changes found in protein levels for these proteins during the first ten minutes after the anaerobic switch.

In addition, using this approach, a number of proteins that have a rapid turnover were also identified. It is unclear whether altered levels of proteins with a short half-life is regulated on the side of the translation or degradation-rate. Of the labile proteins GrcA was found in an earlier study on protein turnover (115), while the half-life of ClpA is also short (264, 265). Upon heat shock, σ^{32} and IbpB were identified as labile proteins, in agreement with other studies (106, 107, 242). While KatG was identified as a labile protein after an anaerobic switch, the functionally related proteins of the AhpC/AhpF complex behaved as proteins with

a short half-life both upon anaerobiosis and heat shock. The functional significance and the mechanisms underlying the rapid turnover of these proteins deserve further investigation.

The increased synthesis-rates of proteins involved in the metabolic adaptations to lack of O_2 raise questions regarding the relative contribution of regulation at the transcription and/or translational level. Recently, relative mRNA levels have been determined at different time points after a switch to a low oxygen environment in a glucose-limited chemostat-culture of *E. coli* using microarrays (267). Comparison of our protein data with these mRNA data (Supplemental Figure 2) clearly shows discrepancies between relative synthesis-rates and changes in mRNA levels for many proteins including those belonging to the glycolysis, PTS and ribosomal proteins, which suggests regulation at the level of translation rather than transcription. It should be stressed that care should be taken when comparing these datasets due to the difference in culture conditions used, and the candidate proteins suggested here to be regulated at the level of translation should be studied in more detail. Clearly, the anaerobic switch seems an interesting experimental system to further study the extent of transcriptional and translational regulation in *E. coli*.

The current study greatly expands prior knowledge of early changes in protein synthesis after a switch to an anaerobic environment. Furthermore, it allows identification of rapidly and slowly degraded proteins on a proteome-wide scale. In combination with genome-wide data on transcript levels our method to determine relative translation rates on a proteomic scale will provide a powerful tool to assess the separate contributions of transcription and translation to the regulation of gene expression.

Supplemental Data—Supplemental Figures and Tables can be found in the addendum section on pages 138-144. Protein identification and quantitation data can be found online at http://www.mcponline.org as supplemental information for ref. (141) and (169).

Experimental Procedures

Synthesis of L-azhal— L-azhal was synthesized from L-Boc-2,4-di-aminobutyric acid (L-Boc-DAB, Chem-Impex, Wood Dale, USA) as described previously (141).

Cell culture— The methionine-auxotrophic *E. coli* strain MTD123 (180) was grown in M9 minimal medium as described before (141). For aerobic to anaerobic-switch experiments two cultures (A and B) were grown aerobically at 37 °C in M9 minimal medium, in order to have a biological replicate. After overnight culture, cells were inoculated at OD_{600} 0.01 and allowed to grow into exponential phase before being harvested at OD_{600} 1.0. Cells were washed at room temperature with complete M9 minimal medium, but lacking methionine (9). Cells were then split and transferred to either a fully anaerobic culture vessel (under nitrogen) or an aerobic culture vessel both containing M9 minimal medium in which the methionine was replaced by 400 mg/l azhal. The four cultures were allowed to resume growth aerobically or anaerobically for 10 minutes before being harvested.

Sample preparation and iTRAQ labelling— Samples were essentially prepared as described before (141), in short: azhal-labelled cells were harvested by centrifugation, pellets were resuspended in lysis buffer and lysed by sonication after which cellular debris was removed by centrifugation. Samples were dialyzed overnight and their protein content was determined. Samples were then subjected to overnight digestion with trypsin, and, for quantitation, digests were labelled with iTRAQ (Applied Biosystems, Toronto, Canada). The digests from the cultures A and B grown under aerobic conditions during pulse-labelling were labelled with iTRAQ 114 and 116 respectively, while digests from the cultures A and B grown anaerobically were labelled with iTRAQ 115 and 117. The four labelled samples were mixed in a 1:1:1:1 (w/w) ratio, based on the protein content of starting material used for trypsin digestion. Excess iTRAQ reagent was removed with a cation-exchange cartridge. Samples were reduced and alkylated before being subjected to COFRADIC.

COFRADIC and mass spectrometric analysis— COFRADIC (213) was applied to enrich azhal-containing peptides, using TCEP to selectively modify target peptides between the primary and secondary chromatographic runs (141). TCEP induces a set of competing reactions in azhal-containing peptides present in primary fractions, i.e., conversion of the azido-group to an amine or hydroxyl group and cleavage of the peptide bond at the C-terminal side of azhal residues (141, 143, 223). The subsequent enrichment is based on a difference in retention-time during the secondary chromatographic runs between TCEP-induced reaction products and the bulk of unmodified peptides that are present in the particular primary fraction subjected to TCEP treatment (141). Three fractions, collected 16 minutes apart in the primary run are pooled and reinjected after TCEP treatment. Shifted fractions of secondary runs enriched in TCEP-induced reaction products from azhal-containing peptides, were analyzed by LC tandem-MS as described in detail previously (141). In addition non-shifted pooled fractions were collected, resuspended in 400 µl of 0.1% TFA with the addition of 150 pmol of human [Glu1]-Fibrinopeptide B (Sigma-Aldrich) for internal calibration. Of these samples, 5 µl was injected and analysed as described for the shifted fractions. Peak lists were generated in Analyst QS 1.1, using the mascot.dll script version 1.6b23, essentially with settings as described on the MASCOT website (http://www.matrixscience.com/help/instruments_analyst.html) with the exception of the precursor mass tolerance for grouping, which was set at 1.0 Da. Assessment of the relative quantity of each protein by analysis of tandem-MS spectra is based on the signal intensities of reporter ions derived from the iTRAQ-moieties of the respective peptides.

Data analysis— Proteins were identified by database searching with the tandem-MS data using the MASCOT search engine version 2.1 (Matrix Science, London, United Kingdom) with parameters as described before (141). In addition a peak list from non-shifted fractions was searched with the following parameters: cleavage after lysine or arginine unless followed by proline, allowing up to one missed cleavage, fixed carbamidomethyl cysteine modification, iTRAQ (K) modification. Variable modification used was the iTRAQ (N-terminal) modification. Peptide mass and MS/MS tolerance was set at 0.1 Da. The significance threshold was set to 0.01 resulting in a threshold score of 28. Multidimensional protein identification technology scoring and "require bold red" were applied with an ion score cut-off of 35 to have all peptide matches identified at a p value of <0.01. MASCOT performed fragment ion searches with the above settings in a local database of the *E. coli* K12 proteome (4328 proteins, 1,381,420 residues, release

11, June 12, 2007, Uniprot consortium). To estimate false positive rates in protein identification we also performed fragment ion searches against a decoy database, which was a shuffled version of the *E. coli* K12 proteome made using the Peakhardt decoy database builder (Medizinisches Proteom Center, Bochum, Germany). False positive identification rates were found to be less than 3.7 percent. The resulting MASCOT data-files of these searches were imported into Quant (245), relative ratios of newly formed proteins were determined by quantitation using the iTRAQ reporter ions of only azhal-containing peptides unique to each protein and stringent criteria for quantitation as described previously (141), while the iTRAQ reporter ions from peptides not containing azhal or methionine were used to quantify changes in protein levels. To ascertain if up- or down-regulation was significant, a double sided Welch's T-test for each protein was performed, using the Welch-Satterthwaite equation (246) to determine the degrees of freedom, to see whether the mean protein ratio differed significantly from the mean ratio obtained for unregulated proteins ($\mu = 1.04$, $\sigma = 0.21$, n=164 for newly synthesized proteins and $\mu = 1.04$, $\sigma = 0.11$, n=302 for protein levels; from ratio of 116/114 and 117/115 reporter ions). Proteins which changed more than 1.5 fold and for which the p-value was adjusted for a false discovery rate of less than 0.05 due to multiple testing (247) were considered to have a significantly altered expression level.

6

General Discussion

Introduction

Different pulse-chase labelling techniques to measure protein synthesis and degradation are available differing in relative strengths and weaknesses within mass spectrometry-based proteomics approaches (as described in *Chapter 1*). In this thesis the development of a pulse-labelling approach using a non-natural amino acid is presented for the model organism *E. coli*. Effects of the label on the physiology of *E. coli* were studied, while selective enrichment of labelled peptides provided the opportunity to use short pulse-labelling times. Application of the azhal pulse-labelling technique in quantitative measurements of changes in newly formed proteins upon changes in growth conditions was successful. In addition, extension of the technique to measure overall protein levels at the same time provided the opportunity to differentiate between stable and labile proteins for those proteins that were significantly up- or down-regulated under the environmental conditions tested. As a consequence pulse-chase labelling using azhal seems poised to take its place among other techniques in the proteome-wide search for post-transcriptional regulation. Here we discuss the strong points and caveats that remain in azhal-based pulse-chase labelling, as compared to pulse-labelling with radiolabels or stable-isotopes as well as other approaches attempting to identify genes that are regulated post-transcriptionally. We also deal with future technical developments that might further improve azhal-based pulse-labelling as well as interesting biological questions that could be answered by its application.

Discussion

Assaying post-transcriptional regulation by ribosomal bound mRNA and azhal pulse-chase labelling— By virtue of genome-wide approaches in measuring transcripts and proteins the importance of post-transcriptional regulation is being further recognized. Analysis of polysomal bound mRNA by microarray (11, 78-80) and deep-sequencing of ribosomal footprints (81) has identified that there is a significant difference between total mRNA and ribosomal bound mRNA in *S. cerevisiae*. These studies indicate that total mRNA levels are an imperfect proxy for the translational status of a transcript and that over- or under-represented ribosome bound mRNA identifies transcripts undergoing translational regulation. These approaches are an additional way of finding genes regulated at the post-transcriptional level and provide data for a vast number of genes as analysis of mRNA using microarray or deep-sequencing benefits from the possibility of amplifying the molecule under study. Analysis of polysome bound mRNA relies on the purification of ribosome bound mRNA, through either affinity purification of ribosomal complexes or sucrose density gradient centrifugation to separate polysomes from other macromolecular complexes. For selective enrichment an affinity tag is introduced into a ribosomal protein (78, 79). The choice of ribosomal protein to be tagged is important as it needs to be recruited into polysomes and its introduction into some ribosomal proteins can lead to growth defects (78). This limits this approach to organisms yielding readily to genetic manipulation. Isolation is a crucial step as biases can

be introduced here resulting in misrepresentation of polysomal complexes, dissociation of ribosomes and mRNA degradation. Affinity purification or density gradient centrifugation to isolate polysomes of the same sample can already give differing results because of this (78). Furthermore, analysis of polysomal bound mRNA has not yet been widely used in prokaryotes. The average half-life of mRNA (~3.7 min) in a prokaryote like *E. coli* (268) is considerably lower than that of mRNA (~20 min) in *S. cerevisiae (*269), which means mRNA degradation is more of a concern.

In comparison, the use of azhal will also be limited to organisms that are either natural methionine-auxotrophes or which can be genetically manipulated to render them auxotrophic, as the K_m/K_{cat} of the methionyl-tRNA-synthetase for azhal favours incorporation of methionine over azhal. However, applicability of azhal pulse-labelling has already been shown in both eukaryotes as well as the prokaryote *E. coli*, and although some proteins might undergo rapid degradation upon azhal-labelling and will thus not be detected, no real bias towards certain proteins is evident for azhal-labelling. An advantage of azhal based pulse-chase labelling is that mass spectrometric detection can be used to quantify protein synthesis-rates (i), total protein levels (ii) and relative protein stability (iii). This not only minimizes differences in technical variation by applying a single analytical technique, it also can address protein half-life, which is an equally important component as translation, in governing changes in total protein levels (*Chapter 1*) over time. Discrepancies between ribosome-bound transcript levels and total protein levels have been suggested to be the result of differences in protein half-life (81). In a recent study in *E. coli* by Taniguchi *et al.* (21) the discrepancy in half-life between cellular mRNA (minutes) and most proteins (hours) was listed as a major cause in discrepancies measured between mRNA levels and protein levels following a perturbation. As such, although the mRNA complement measured reflects the pool available for translation at that moment, as does the rate of new protein formation measured by azhal incorporation, the protein amounts in the cell are a blend of proteins expressed at that particular time point and surviving pre-existing proteins in the cell. Therefore a more complete description of the proteome than that of mere changes in protein levels is necessary, as changes in protein levels are a result of the interplay of protein synthesis and degradation. As demonstrated in *Chapters 4 and 5* pulse-labelling with azhal can measure relative synthesis-rates in conjunction with protein levels and give a measure of protein stability under non-steady state conditions. All in all, azhal-labelling gives an even more direct measure of new protein formation than ribosomal association of transcripts. However, for identifying genes which are regulated through differential translation rates of transcripts, ribosomal occupancy is an equally valid measurement. As such, either approach can be applied as a validation of forms of post-transcriptional regulation identified and test for possible biases or caveats within the other approaches to identify translational regulation.

Pulse-chase labelling with azhal, in relation to radiolabelling and stable-isotopes— Direct comparison of the different approaches is not straightforward as studies have been performed in different model organisms, with different growth rates, addressing both synthesis and degradation. Azhal as a label seems applicable in a variety of organisms as is the case for both stable-isotope and radio-isotopes. However results obtained for *B. subtilis* and *S. cerevisiae* (*Chapter 2*) show that use of azhal is more limited and labelling and growth needs to be tested for each prospective organism. Furthermore, the much lower K_m/K_{cat} of the methionyl-tRNA-synthetase for azhal makes it highly preferable to use auxotrophic organisms, to preclude the presence of endogenous methionine. This requires extensive washing to remove methionine which can introduce limitations on time-series under experimental conditions which cannot be easily maintained during wash steps (e.g. keeping cells anaerobic). Methionine limited continues-culture could resolve the need for washing as methionine in the culture vessel would be virtually absent. However, this approach is only applicable to micro-organisms that can be cultured in a chemostat in a chemically defined medium.

With respect to applying pulse-chase labelling in a proteomics study using mass spectrometry both azhal as well as stable-isotope labelling are directly comparable. They offer the possibility to extract both the protein identity as well as synthesis and degradation-rates from the mass spectral data directly. This can prevent some of the drawbacks related to the use of two dimensional gel electrophoresis in combination with radiolabelling and mass spectrometry as described in *Chapter 1*. Temporal resolution of radiolabelling is such that short pulse-labelling times can record rapid changes in protein synthesis following a change in environmental conditions. In *E. coli* for example, changes for a number of proteins upon a change in growth temperature or an anaerobic switch, have been quantified previously (122, 124). Stable-isotopes are less suited for these types of studies as the considerable degree of labelling required (135) limits their applicability on these short time-scales. Labelling studies that employ stable-isotopes typically look at synthesis and degradation-rates in a steady state (e.g. exponential phase) of growth (112-121). Under these conditions degradation (taking into account growth- or dilution-rate) should be equal to synthesis-rate, and it is feasible to compare different steady states. We have shown in *Chapters 4 and 5* that with azhal-based pulse-labelling initial changes in formation of new proteins following an environmental switch (growth temperature and anaerobic switch) can be quantified and are in good accordance with radiolabelling data. Although temporal resolution is still somewhat lower for azhal-labelling experiments, the number of individual proteins that are monitored simultaneously is *far* greater. In addition, future developments in both COFRADIC and the enrichment technologies (see below) can push the temporal resolution alongside that of radiolabelling without compromising the number of proteins that can be quantified.

Degradation-rates can play an equally important role as synthesis-rates in the regulation of cellular protein levels, and we have shown that it is feasible to discern stable from labile proteins by quantifying protein levels in addition to measuring newly synthesized proteins on the same time-scale. Half-lives can only be estimated for proteins that show a significant

change in the amount of newly synthesized species under the growth conditions compared, which is a limitation. A sudden change in the level of a particular protein without any change in amount of newly formed copies of this protein would also be a good indication of a change in degradation-rate, but this was not observed in the datasets acquired thus far. While the use of stable-isotopes is less suitable to determine changes in translation rates by pulse-labelling, stable-isotope-labelled amino acids are suitable to determine protein turnover. For instance, half-lives of over 600 proteins were measured in an adenocarcinoma cell line (116) with this approach. It remains to be seen, however, if stable-isotope labelling can also pick up transient changes in protein degradation-rate.

Improvements of azide-enrichment methods— The enrichment of azhal-containing peptides or proteins is essential for the sensitivity and temporal resolution of the azhal-labelling approach. We employ a peptide centric enrichment scheme. Azhal-containing peptides are enriched by a TCEP-induced retention-time shift of labelled peptides between two reversed-phase chromatographic separations as described in *Chapter 3*. Both by the number of azhal-containing proteins identified (over 500) after a short pulse-labelling (15 minutes) as well as its application in a quantitative proteomics approach in *Chapters 4 and 5* we show the sensitivity and robustness of the COFRADIC enrichment technique.

However, in its current form COFRADIC enrichment is quite labour-intensive, and requires considerable runtime on an LC-system. Currently a single enrichment takes 24 hours of continuous LC-runtime on two instruments which need manual attention for injecting samples as well as removing and pooling collected fractions. This is the reason why we employ iTRAQ (four labels) as our quantitative technique of choice, as its multiplexing capabilities lessen the number of enrichments necessary per sample set. Further development of COFRADIC enrichment could entail the setup of a robotic system coupled to the LC configurations to automate injection of samples, TCEP reactions, and the removal and pooling of collected fractions to enable a truly 24 hour workload for the LC-systems employed. Use of the eight-label version of iTRAQ could further reduce the number of enrichments needed per sample set, reducing the amount of LC-runtime required.

Using the reaction of TCEP with azhal-containing peptides, we do not expect that pulse-labelling times much shorter than 10 minutes will result in de detection of large numbers of proteins. This is due to peak broadening of the main chromatographic peaks that contain the bulk of unlabeled material and the, roughly equimolar, formation of no less than four different reaction products. This reduces the signal intensity measured per product by a factor of three (azhal-containing peptide can enter one of three different reactions cascades). Employing a reaction against azhal-containing peptides that only leads to the formation of a single product would therefore increase the sensitivity by a factor of three. This would enable reduction of pulse-labelling times without decreasing the number of proteins identified and quantified. There are different reactions against azides described in *Chapter 1* which could be employed for this purpose. Side-reactions however might occur as we experienced for the Cu^{I} catalyzed

(3+2) cyclo-addition and the addition of a cyclo-octyne introduces a hydrophobic group which might result in peptide losses due to solubility problems. Consequently, replacing the TCEP would require the synthesis of more hydrophilic octynes (Figure 1) such as described by Sletten *et al.* (157) to minimize peptide loss during enrichment. However, if other types of cyclo-octynes (Figure 1b and c) are considered, care has to be taken that these do not form fragments upon collision-induced dissociation that would make interpretation of tandem mass spectra with proteome database search engines difficult as described by Nessen *et al.* (173) for the mono-fluorinated cyclo-octyne and azacyclo-octyne of Sletten *et al.* (157). Another possibility is the Staudinger ligation using phosphines (161) that, in contrast to TCEP, favour ligation over cleavage and reduction of azhal-containing peptides. Azide-reactive moieties such as electron deficient alkynes (60) and oxanorbornadienes (270) can also be considered as alternatives for TCEP. However, these also have their own drawbacks such as low reactivity (electron deficient alkynes) or side-reactions (oxanorbornadienes) (173). As such, replacing TCEP as a reactive agent might be less straightforward than it seems, but the increase in sensitivity and potential reduction of labelling times, could make the effort worthwhile.

In our COFRADIC approach to isolate azhal-labelled peptides we combine three primary fractions out of a total of 48 for each secondary run (*Chapter 3*). This implies that the di-aminobutyrate- and homoserine-containing peptides derived from the azhal-containing species after TCEP treatment are likewise eluting in three somewhat broadened fractions, given the relatively narrow time window of shift times in reversed phase chromatography of TCEP-induced reaction products with respect to their parent compounds. Identification of newly synthesized proteins after azhal pulse-labelling is mainly based on MS/MS of these di-aminobutyrate- and homoserine-containing peptides (*Chapter 3*). After pooling the shifted fractions in COFRADIC, the three-peak elution pattern will be roughly reflected

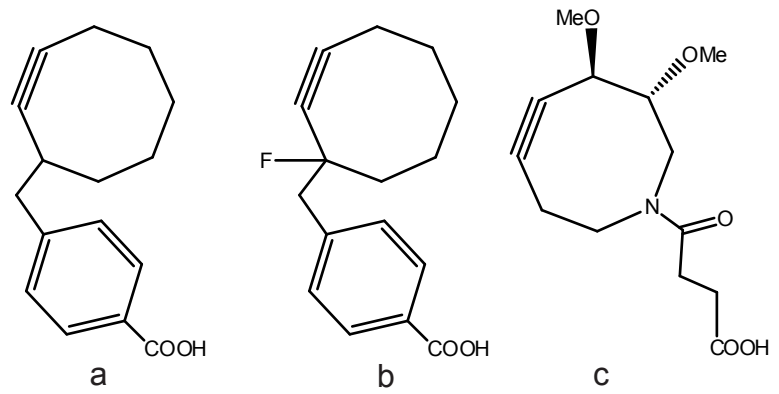


FIGURE 1. **Different cyclo-octynes used in azide-reactive affinity resins.** The cyclo-octyne (**a**) is successfully used to enrich azide-containing peptides with ARCO-resin as described by Nessen *et al.* (170), whereas use of the mono-fluorinated cyclo-octyne (**b**) or the more hydrophilic azacyclo-octyne (**c**) in azide-reactive affinity resins generated fragments during CID which significantly decreased proteins identified by the proteome database search engine MASCOT (173).

in the reversed phase LC fractionation preceding mass spectrometric analysis, since this LC step is carried out under similar conditions (low pH, acetonitrile gradient) as used in COFRADIC reversed phase chromatography. More orthogonality in separation between the COFRADIC-enrichment and LC-MS analysis could be achieved by increasing the pH of the LC solvents employed during enrichment (271).The retention-time shift of di-aminobutyrate and homoserine reaction products under these LC conditions should be tested to ascertain if they are still sufficient to separate them from the bulk of unlabeled peptides. The higher orthogonality between the separations during enrichment and analysis may increase the number of proteins identified and quantified per LC-MS run.

An alternate approach towards enrichment of azhal-containing molecules, is one using affinity purification as presented in *Chapter 1*. Although still in the proof of principle phase of development, affinity enrichment offers several potential advantages over the COFRADIC enrichment scheme. If enough starting material can be acquired, the reduction of labelling times should not pose a limitation for an affinity enrichment approach. Furthermore, it can be used in both a peptide- as well as a protein-based approach to enrichment. The latter offers the potential to use multiple stable-isotopes in addition to azhal during pulse-labelling in order to expand the set of peptides that can yield quantitative information about protein synthesis or degradation-rates. Finally an affinity approach is more amenable to the workup of multiple samples at once, increasing the throughput compared to COFRADIC.

Dieterich *et al.* (167, 168) have demonstrated the application of affinity tagging proteins via copper-catalysed azide-alkyne chemistry and subsequent enrichment through the interaction of the biotin tag with an avidin column after a 2 hour pulse-labelling period in HEK-cells. Here a second pulse-label, i.e. $^{2}H_{10}$-leucine, was employed and detection of this pulse-label accounted for the identification of most labelled peptides. All but one derivatized azhal-containing peptide escaped detection, and non-

derivatized azhal-containing peptides were a minority amongst labelled peptides detected (*Chapter 1*). This can be attributed to the on-bead digestion which might not have been efficient in cleaving the linker of the affinity tag, but did liberate tryptic peptides from captured proteins. Azhal basically functions as an affinity handle only to enrich newly formed proteins here, while $^{2}H_{10}$-leucine is used to identify them. Future developments in this approach could entail a different set of cleavable linkers, used in a cleavable alkyne-biotin affinity tag (272). A more efficient liberation of azhal-containing peptides can expand the coverage of peptides directly linked to labelling and new protein formation. This is of interest for quantitative experiments because only labelled peptides will give accurate quantitative data about new protein formation, whereas levels of non-labelled peptides can be influenced by aspecific binding of non-labelled proteins. However concerns remain regarding possible side-reaction(s) of the copper catalysed (3+2) cyclo-addition mentioned in *Chapter 1*.

In tandem with the COFRADIC approach we developed an affinity label both for the enrichment of azhal-containing peptides, and azide-containing cross-linkers as described in *Chapter 1*. Based on strain-promoted (3+2) cyclo-addition between azides and cyclo-octynes,

this method has been used for the enrichment of azhal labelled peptides in *E. coli* following extended labelling times (170). It shows promise in reducing these labelling times down to levels common for radiolabelling approaches (173). In contrast to COFRADIC, which is peptide-based by nature, the same affinity label should also enable enrichment of labelled proteins. This has several advantages with respect to co-labelling, as described in the above.

Some drawbacks of the current peptide-centric approach using the ARCO-resin are both the hydrophobicity of the cyclo-octyne, coupled to the azhal-peptide following release, and the disulfide bond as a cleavable group within the linker. The first results in peptide losses due to solubility problems and also requires adaptation of gradients to provide better separation of these more hydrophobic peptides. Use of less hydrophobic cyclo-octynes (157) may improve solubility of peptides, but use of alternate cyclo-octynes should not hamper identification of peptides by proteome-database search engines, as described for replacing TCEP reactions by strain-promoted (3+2) cyclo-addition. The choice of a disulfide bond as the cleavable group in the linker can cause unwanted background due to disulfide exchange between cysteine-containing peptides and the disulfide linker (170). Rigorous reduction and alkylation, could alleviate this problem, but extended reduction/alkylation protocols could affect azhal-containing peptides as well, as the azide is susceptible to conversion by common reducing agents (143). It is clear that the different affinity approaches need further development to match the utility of COFRADIC enrichment in quantitative proteomics approaches to study biologically relevant questions. Efforts towards improving these approaches can be well worth it though, as a successful affinity enrichment of azhal-containing peptides or proteins may confer higher temporal resolution to azhal pulse-labelling competitive with that of radiolabelling approaches.

Future applications of azhal pulse-chase labelling— Quantitation of relative synthesis-rates has been shown in *E. coli* by the use of azhal as a pulse-label in *Chapter 5*. Further developments in enrichment procedures can increase temporal resolution as described above and enable direct quantitation of relative synthesis-rates for more labile proteins as well. Comparison of relative amounts of new protein formation after anaerobic shift (169) and heat shock (141), respectively, showed discrepancies between changes in mRNA levels (from literature) and formation rates of new proteins (*Chapter 4 and 5*). Although care should be taken when datasets are compared due to differences in culture conditions, and candidate proteins found here should be studied in more detail, this already shows the utility of comparing new protein synthesis and transcript level changes. Future experiments with azhal pulse-labelling in *E. coli* should be accompanied by microarray or deep-sequencing quantitation of mRNA level changes in the *same* experiment. Identification of genes regulated at the translational level by azhal pulse-labelling can also be further validated by measurements of polysome bound mRNA in conjunction with total mRNA and total protein levels as described above. This will enable more unambiguous identification of candidate proteins undergoing translational regulation. This in turn can be the starting point of a myriad of follow-up studies

into the molecular mechanisms underlying post-transcriptional regulation of these genes.

We already postulated that the aerobic to anaerobic switch in *E. coli* might be an interesting system in which to study post-transcriptional regulation based on the results obtained so far (169). In this context it is interesting to note that the Sm-like protein Hfq, which functions as a global translational regulator, has been found associated, among many other mRNAs, with nearly all mRNAs encoding glycolytic enzymes and proteins of the phosphoenolpyruvate-dependent sugar phosphotransferase system (PTS) in *Salmonella enterica* (273). Hfq is an mRNA chaperone and has been found to mediate the action of many sRNAs in *E. coli* **(***Chapter 4***)**. These sRNAs affect translation in various ways through imperfect base-pairing with their mRNA targets and a large number of them have been identified in *E. coli* (*Chapter 1* and *4*). The Hfq-mediated posttranscriptional regulation by the sRNA *sgrS* of *ptsG* expression (one of the PTS genes) has been studied before (274, 275). As such it would be of interest to study the effects that deletion of Hfq, and/or different small regulatory RNAs, have on transcription and translation following an anaerobic switch using azhal-labelling and microarray studies. Such experiments could shed new light upon sRNA regulation of the glycolytic enzymes in *E. coli* during changes in oxygen availability.

Another possible expansion is the application of quantitative azhal pulse-labelling in other organisms than *E. coli*. A number of mammalian cell types (163, 166-168) and even *D. melanogaster* (171) seem to be amenable to azhal-labelling, though the lack of growth on azhal found for *B. subtilis* and *S. cerevisiae* shows that this label is not universally applicable. The setup of azhal-labelling in any prospective organism should be similar to that in *E. coli*, irrespective of whether COFRADIC or another enrichment approach is chosen. Growth, kinetics of labelling and toxicity should first be tested. Results obtained by azhal pulse-labelling should be compared to those obtained by other pulse-labelling approaches or polysome bound mRNA measurements in the organism. The expansion of a quantitative azhal pulse-labelling approach to higher organisms is of interest, as the role of translational regulation in these organisms is thought to be more prevalent and extensive than in prokaryotes.

Further development of the ‘chase-type’ of experiment using azhal-labelling is limited by the growth arrest observed in *E. coli*, although this may not occur in other organisms. Without full labelling, estimation of protein half-life in the case of rapid turn-over is difficult, although probably still feasible for proteins with a longer half-life. Care should be taken, however, when setting up chase-experiments with azhal. Proteins that do not fold correctly if azhal is incorporated (tentatively the case for LacZ as described in *Chapter 2*) could show an increased turnover rate due to azhal-labelling. This would erroneously identify such proteins as labile. This is less of a concern for measuring new protein formation, as the rapidly degraded proteins would simply escape detection. In the case of relative quantitation of changes in protein half-lives between different growth conditions grown in the presence of azhal, this could also be less of a problem as the rate of decay of these misfolded proteins would be expected to be equal under most conditions, if not regulated otherwise. Although

concerns about artefacts remain, the proteome-wide measurement of changes in degradation-rate during non-steady state conditions is of interest, as transient changes in protein stability can be important for regulation as described for σ^{32} in *E. coli* in *Chapter 1*.

Conclusion and outlook— All things considered, azhal pulse-labelling in *E. coli* shows the added value of determining protein synthesis-rates and half-lives, in addition to total protein levels, during transient changes induced by environmental conditions. Future applications in *E. coli* and other organisms in conjunction with genome-wide measurements of transcript levels could identify genes that are subject to post-transcriptional regulation as protein synthesis-rate should be closely linked to transcript level if translational regulation does not occur. Determination of the half-lives and their regulation in conjunction with synthesis-rates can aid in elucidating whether post-transcriptional regulation occurs via synthesis, degradation, or both on a proteome-wide scale. Together with reduction of pulse-labelling times required for the simultaneous analysis of hundreds of proteins this will add a new layer to the analysis of cellular proteome dynamics.

7

Summary & Samenvatting

Summary

Understanding of the highly complex regulation of cellular physiology has greatly benefited from various genome-wide approaches measuring transcripts, proteins and metabolites. Genome-wide studies have shown that for a large number of proteins, transcript levels do not directly reflect cellular protein levels, as one might dogmatically have expected. Protein expression can be regulated at various points going from transcript to protein. Cellular protein levels are governed by protein half-life as well. *Chapter 1* gives an overview of post-transcriptional regulation that affects either translation or degradation of proteins. Determination of protein synthesis and degradation for many proteins simultaneously, will be invaluable to pinpoint where post-transcriptional regulation takes place. Currently, genome-wide determination of protein synthesis as well as half-life is difficult. Especially when transient changes in synthesis or degradation need to be detected during cellular adaptation. Classically, protein synthesis and degradation are determined by pulse-chase labelling using radio-isotopes. Although this approach offers high temporal resolution, it is difficult to integrate within a mass spectrometry based proteomics approach. However, the use of stable-isotopes in stead of radio-isotopes, although supremely compatible with mass spectrometry, offers only limited temporal resolution. The latter is caused by the bulk of unlabelled protein, which obscures labelled species, especially if labelling times are short.

Non-natural amino acids are an alternative to stable-isotopes. They combine mass spectrometric compatibility with (much) higher temporal resolution. The methionine analogue azhal is such a pulse-label that has been shown to be incorporated into proteins by *E. coli,* mammalian cell lines and cultured insect cells. Non-natural amino acids achieve higher temporal resolution in the face of short labelling times through selective enrichment of labelled proteins or peptides from the unlabelled background. The azhal-labelled molecules are enriched by covalent attachment to affinity-resins through various forms of ‘click chemistry’ directed against the azide-moiety of azhal. This enables sensitive mass spectrometric detection of low abundant azhal-labelled peptides or proteins.

Chapter 2 describes the physiological response of two prokaryotic model organisms (*E. coli* and *B. subtilis*) to azhal. Furthermore, the effects of azhal incorporation on protein structure and function are investigated in different recombinant proteins. *E. coli* grows equally well on azhal as on methionine during the first 30 minutes upon substitution of the latter, after which growth arrest gradually sets in. In contrast, *B. subtilis* grown on the analogue has an initial lag phase and a considerably lower growth rate. Three photo-active proteins (PYP, AppA and YtvA) labelled with azhal do not show evidence of aberrant folding. Upon illumination however, these proteins display somewhat altered recovery rates from signalling to ground state, compared to their methionine containing counterparts. Azhal labelled LacZ, however, cannot be produced, probably due to misfolding and rapid degradation of this protein caused by azhal incorporation.

In *Chapter 3* an alternative approach to affinity enrichment of azhal-labelled peptides is presented. This enrichment is based on a change in retention-time induced by the selective reaction of the azide-moiety with tris(2-carboxy-ethyl)-phosphine. This reaction induces no less than four different reaction products in azhal-containing peptides or proteins, three of which are described here for the first time. Selectively modified peptides enriched by the retention-time shift are subsequently identified by tandem-MS. Following a pulse-labelling period of only 15 minutes, 527 proteins representative of all major Gene Ontology categories are identified in *E. coli* using this enrichment approach.

Chapter 4 describes the quantitative application of the enrichment approach. iTRAQ is used to quantitatively compare azhal-labelled peptides between growth conditions, representing relative changes in new protein formation. The initial phase after a change in growth temperature from 37 ºC to 44 ºC in *E. coli* is studied. Measurement of the relative amounts of 344 proteins newly synthesized in 15 minutes upon a switch in growth temperature showed that nearly 20% in- or decreased more than two-fold. Most regulated proteins detected have functions as chaperones or proteases, in accordance with this change in growth conditions, while the changes in new protein formation are highly similar to those found in earlier radiolabelling studies. In addition, the analytical strategy is extended to determine changes in total protein levels on the same time-scale, using quantitative data from the non-shifted peptides not containing azhal. Collation of changes in protein levels with changes in newly synthesized proteins enables identification of ‘stable’ and ‘labile’ proteins. The vast majority of proteins were found to be stable, only a subset of 5 proteins having a higher turnover rate under these growth conditions.

The extended strategy is also used in *Chapter 5* to determine average relative translation rates for 10 minutes immediately following a switch from aerobiosis to anaerobiosis. The majority of proteins with increased synthesis-rates upon an anaerobic switch are involved in glycolysis and pathways aimed at preventing glycolysis grinding to a halt by a cellular redox-imbalance. Newly formed proteins, quantified following heat shock (*Chapter 4*) and a switch to anaerobiosis (*Chapter 5*), are also compared with microarray data from literature obtained under similar conditions. Surprisingly, this reveals -for the first time- that regulation following a temperature increase is predominantly transcriptional, whereas for a substantial number of proteins translational regulation seems to be used upon sudden anaerobiosis. This illustrates the utility of azhal-based pulse-labelling to probe translational regulation.

Chapter 6 deals with strong points as well as remaining caveats in pulse-chase labelling using azhal, as presented in the previous chapters. The approach is compared to pulse-labelling with radio- or stable-isotopes and other approaches attempting to identify

genes that are regulated post transcription. Future technical developments in both affinity enrichment as well as chromatographic enrichment are introduced. Together with the further reduction in labelling times, these improvements could provide analyses of synthesis-rate and stability of hundreds of proteins, truly adding new layers to the analysis of cellular proteome dynamics.

Samenvatting

Ons inzicht in de complexe regulering van cellulaire fysiologie heeft veel baat gehad bij verschillende genoomwijde analyses van transcripten, eiwitten en metabolieten. Genoomwijde studies hebben aangetoond dat transcript niveaus niet altijd direct overeenkomen met cellulaire eiwitniveaus, zoals menigeen misschien had verwacht. Eiwit expressie kan afgezien van regulatie middels transcriptie ook gereguleerd worden gedurende de translatie van een transcript naar een eiwit. Afgezien van de translationele controle worden cellulaire eiwitniveaus ook door de halfwaardetijd van eiwitten bepaald. *Hoofdstuk 1* geeft een overzicht van post-transcriptionele regulatie, welke eiwitaanmaak en afbraak beïnvloedt. Het bepalen van eiwitsynthese en afbraak voor grote aantallen eiwitten tegelijkertijd zal van onschatbare waarde zijn om te kijken waar en hoe post-transcriptionele regulatie optreedt. Op het moment is het genoomwijd bepalen van eiwitsynthese snelheden en halfwaardetijden lastig. Dit in het bijzonder wanneer kortstondige veranderingen in synthese of afbraak gedetecteerd dienen te worden gedurende cellulaire adaptatie. Over het algemeen wordt eiwitsynthese en –afbraak bepaald door middel van 'pulse-chase' labelen, m.b.v. radio-isotopen. Hoewel deze aanpak met radio-isotopen een hoge temporele resolutie biedt, is het lastig deze te gebruiken met een op massaspectrometrie gebaseerde proteomics aanpak. Helaas heeft het gebruik van stabiele-isotopen (in plaats van radio-isotopen), hoewel optimaal voor massaspectrometrie, een beperkte temporele resolutie, hetgeen wordt veroorzaakt door de bulk van ongelabelde eiwitten. De detectie van gelabelde eiwitten wordt speciaal bemoeilijkt wanneer korte pulse-label tijden gebruikt worden.

Niet natuurlijke aminozuren vormen een alternatief voor stabiele-isotopen. Zij geven compatibiliteit met massaspectrometrie en hoge temporele resolutie. De methionine analoog azhal is zo een pulse-label, waarvoor reeds incorporatie in eiwitten door *E. coli*, zoogdiercellijnen en gekweekte insectcellen is aangetoond. Niet natuurlijke aminozuren kunnen een hogere temporele resolutie bereiken ondanks korte labeltijden door selectieve verrijking van gelabelde eiwitten of peptiden uit de ongelabelde achtergrond. Azhal gelabelde moleculen worden verrijkt door middel van covalente binding aan affiniteits harsen door 'click chemie' gericht op de azide groep van azhal. Dit vergemakkelijkt gevoelige massaspectrometrische detectie van laag abundante azhal gelabelde peptiden of eiwitten.

Hoofdstuk 2 beschrijft the fysiologische respons van twee prokaryote model organismen (*E. coli* en *B. subtilis*) op azhal. Verder worden de effecten van inbouw van azhal in verscheidene recombinante eiwitten onderzocht. *E. coli* groeit even goed op azhal als methionine gedurende de eerste 30 minuten na vervanging, waarna de groeisnelheid langzaam afneemt. Daarentegen vertoont *B. subtilis* een initiële 'lag' fase en een duidelijk lagere groeisnelheid. Drie fotoactieve eiwitten (PYP, AppA en YtvA) gelabeld met azhal lijken normaal te vouwen. Echter, na belichting herstellen deze eiwitten met een ietwat veranderde snelheid vanuit aangeslagen toestand naar grondtoestand in vergelijking met hun

methionine bevattende tegenhangers. In tegenstelling tot de foto-eiwitten kan azhal gelabeld LacZ niet gemaakt worden, waarschijnlijk door verkeerde vouwing en versnelde afbraak van het eiwit als gevolg van azhal inbouw.

In *Hoofdstuk 3* wordt een alternatieve aanpak voor affiniteitzuivering gepresenteerd. Deze verrijking is gebaseerd op veranderingen in retentie tijd na een selectieve reactie van de azide groep met tris(2-carboxy-ethyl)-phosphine. Deze reactie geeft maar liefst vier verschillende reactie producten in azhal bevattende peptiden of eiwitten, waarvan drie hier voor het eerst beschreven worden. Selectief gemodificeerde peptiden verrijkt door de retentietijd verschuiving worden vervolgens geïdentificeerd door tandem-MS. Op deze manier worden 527 *E. coli* eiwitten representatief voor alle belangrijke Gen Ontologie categorieën geïdentificeerd met maar 15 minuten azhal labelen.

Hoofdstuk 4 beschrijft de kwantitatieve toepassing van de chromatografische verrijkingsmethode. iTRAQ wordt gebruikt om azhal gelabelde peptiden uit verschillende groeicondities te kwantificeren, wat veranderingen in nieuwe eiwit vorming weergeeft. De initiële fase na een verandering in groeitemperatuur van 37 °C naar 44 °C in *E. coli* wordt bestudeerd. Meting van de relatieve hoeveelheden van 344 nieuw gemaakte eiwitten in de15 minuten volgend op de verandering in groeitemperatuur toont dat bijna 20% toe- of afneemt met een factor van twee of meer. De meeste gereguleerde eiwitten welke gevonden worden zijn chaparones of proteases zoals verwacht bij deze verandering in groeicondities. Bovendien waren de veranderingen in vorming van nieuwe eiwitten in overeenstemming met eerdere studies uitgevoerd met radio-isotopen. Daarnaast is de analytische aanpak uitgebreid, om ook veranderingen in totale eiwitniveaus op dezelfde tijdsschaal te bepalen. Hiervoor worden de kwantitatieve data van de niet verschoven
-azhalvrije- bevattende peptiden gebruikt. Vergelijking van de veranderingen in zowel eiwit niveaus als synthese snelheden maakt de identificatie van 'stabiele' en 'labiele' eiwitten mogelijk. De grote meerderheid van de gemeten eiwitten vertegenwoordigen stabiele eiwitten, slechts 5 eiwitten hadden een hoge afbraaksnelheid onder de gebruikte groeiomstandigheden.

De uitgebreide aanpak wordt ook in *Hoofdstuk 5* gebruikt om de relatieve translatiesnelheid te bepalen gedurende 10 minuten na een verandering van aeroob naar anaeroob milieu. De meerderheid van eiwitten met verhoogde synthesesnelheden onder deze condities waren direct betrokken bij glycolyse of bij cellulaire paden die stilstand van de glycolyse door redox onbalans moeten voorkomen. Nieuw gevormde eiwitten die werden gekwantificeerd volgend op een hitteschok (*Hoofdstuk 4*) dan wel na plotselinge anaerobiose (*Hoofdstuk 5*) zijn ook vergeleken met microarray studies uitgevoerd onder vergelijkbare omstandigheden. Hieruit bleek -voor de eerste maal- dat regulatie volgend op een temperatuurstoename voornamelijk transcriptioneel is, terwijl na zuurstof depletie translationele regulatie bij een aanzienlijk aantal eiwitten lijkt op te treden. Dit toont het nut

van pulse-labelen met azhal om translationele regulatie op te sporen.

Hoofdstuk 6 behandelt de sterke en zwakke punten van pulse-labelen met azhal zoals beschreven in de voorgaande hoofdstukken. De aanpak wordt vergeleken met pulse-labelen met zowel radio- als stabiele-isotopen en andere technieken die genen proberen te identificeren welke post-transcriptionele regulatie ondergaan. Toekomstige technische ontwikkelingen in zowel affiniteitzuivering als chromatografische verrijking worden geïntroduceerd. Samengaand met voortgaande reductie van label tijden door deze verbeteringen, zal de analyse van synthese snelheid en stabiliteit van honderden eiwitten m.b.v. azhal-labeling de studie van cellulaire proteoom dynamiek verder helpen verdiepen.

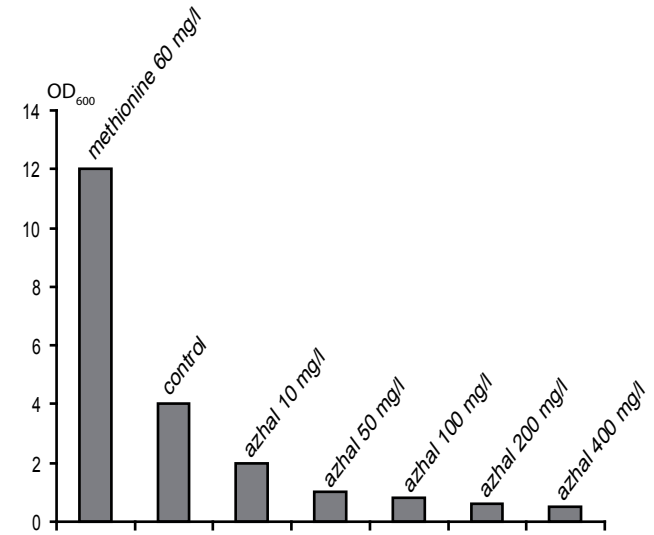
A

Addenda

Addendum Chapter 2

Experimental Procedures

Cell culture— S. cerevisiae strain BY4741 was grown in YPD medium aerobically at 30 °C For growth experiments cells were transferred to minimal medium (276) containing 60 mg/l of all natural amino acids. Cells were inoculated at OD_{600} 0.1 and allowed to grow into exponential phase before being harvested at OD_{600} 1.0, by spinning down the cells for 10 minutes at 4500 rpm and 4 °C. Cells were then washed twice by resuspending the cell pellet in sterile minimal medium without additives (followed by centrifugation) to eliminate traces of methionine. After washing, cells were transferred (at OD_{600} 0.1) to minimal medium in which the methionine was replaced by azhal and cells were allowed to resume growth aerobically at 30 °C. Cells were allowed to grow overnight with varying methionine or azhal conditions as indicated, before the OD_{600} reached overnight was measured.



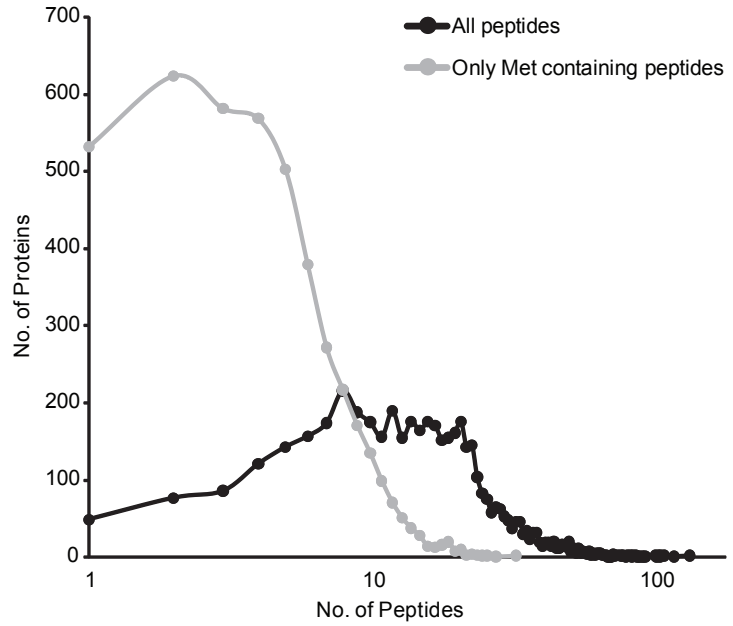
Supplemental Fig.1. I**nhibition of growth of *S. cerevisiae* by addition of azhal.** Yield of overnight growth of *S. cerevisiae* is compared for different concentrations of azhal in minimal medium compared to methionine or control (neither azhal or methionine added). Clearly addition of azhal to minimal medium inhibits growth yield below that of control and increase of azhal concentration results in a dose dependent lowering of growth yield after overnight culture

Addendum Chapter 4

Experimental Procedures

In silico digests and residue content determinations— To determine the number of proteins that can theoretically be covered by methionine containing peptides and the number of proteins that are available within a specific mass window, an *E. coli* K12 proteome database (4691 protein entries, release 2010_05, April 20, 2010, Uniprot consortium) was digested in-silico using the DBTOOLKIT version 1.4 (277). Settings: trypsin digestion, allowing 0 missed cleavages and setting the mass window for peptides from 500 to 3800 amu. Missed cleavages were not allowed as examination of MASCOT identifications of various large datasets revealed the number of assignments involving missed cleavages to be only ~2%, while the mass window was chosen because masses of MASCOT identified peptides from the QSTAR-XL fell within this range. Subsequently the digest-database was exported in FASTA-format and a filtered digest-database with only methionine containing peptides was created, using DBTOOLKIT, and exported as well.

Both filtered and unfiltered peptide sets were imported for further analysis in Excel (2007, Microsoft corporation, Redmond, USA) containing the freely available ASAP-utilities add-on (A must in every office B.V., Zwolle, The Netherlands). Because all N-terminal peptides contain a methionine and a substantial part of proteins in *E. coli* are processed modifying the N-terminus often resulting in removal of the N-terminal methionine residue, we first removed all N-terminal peptides from both digest databases created in DBTOOLKIT. Subsequently by using the pivot table function to count the number of peptides per protein and the number of proteins identified by a certain number of peptides a distribution for both sets was created. In addition the total percentage of entries theoretically covered by de different peptide sets were calculated as well by counting the number of unique entries identified by the two digest databases. To determine the methionine residue contents of identified proteins, their sequences were retrieved from Uniprot in FASTA-format (www.uniprot.org) and imported into Excel. The total number of residues and the number of methionine content were counted using the ASAPCOUNTCHAR feature of ASAP-utilities, these were used to calculate the average methionine content of the identified proteins.



Supplemental Fig.1. **Number of peptides per protein after an in silico digest of the *E. coli* proteome database.** The number of proteins for which a certain number of peptides are found within the mass range of 500-3800 amu after an in-silico digest, not allowing missed cleavages, of the *E. coli* proteome database are shown. The light grey line shows the distribution for all peptides, while the dark grey line shows the distribution for peptides containing methionine residues only. From the 4691 protein entries in the *E. coli* database 4677 (99.7%) are theoretically identified when considering all peptides while 4369 (93.1%) are theoretically identified when only considering peptides containing a methionine residue. Obviously, although the total number of proteins covered is not much reduced when only employing methionine containing peptides, the number of peptides theoretically detectable per protein declines sharply as the total number of methionine containing peptides (22585) is only 25.6% of the total number of tryptic peptides (88287).

Addendum Chapter 5

Kinetic Model

To extract a measure for the protein translation rate from our data set we constructed a kinetic model that predicts the number of newly synthesized proteins in the course of an azhal pulse-labelling experiment. Pulse-labelling is carried out in the exponential growth phase of *E. coli* cells in a culture flask. We assume that during the exponential growth phase the number of proteins increases proportional to the total of cellular mass and that protein synthesis (k_{syn}) and degradation (k_{deg}) rates are first-order with the number of proteins as described by Mosteller and Goldstein (278), in equation (a) where F(t) represents the time dependent increase in cellular mass.

$$\frac{\partial P_{total}}{\partial t} = k_{syn}F(t) - k_{\deg}P_{total} \qquad (a)$$

At time point t is zero minutes we have only unlabeled proteins P_{old}. Post azhal-labelling we have at t = 10 minutes the partly degraded old protein population P_{old} and we have newly synthesized azhal-labelled proteins P_{new}. The degradation-rate of the old unlabeled population P_{old} is given by differential equation (b).

$$\frac{\partial P_{old}}{\partial t} = -k_{\deg}P_{old} \qquad (b)$$

During labelling azhal-containing new proteins P_{new} are formed with a first-order rate constant k_{syn} and degrade with rate constant k_{deg}. We assume that the rate constant k_{deg} is not influenced by azhal replacing methionine. From literature it is known that in vivo degradation-rates of some amino acid analogue-labelled proteins can differ from their natural counterparts (279, 280). However compared to these analogues, *E. coli* can sustain normal growth longer with azhal. In addition, there is evidence for unperturbed protein processing of the N-terminal azhal residue in *E. coli* (175) and for normal localization and folding of azhal-labelled proteins (148-151). Most importantly, the relative change in average synthesis-rate upon anaerobiosis found by Smith *et al.* (124) using radiolabelling for PflB is similar to that found by azhal-labelling, and similar results were found by azhal-labelling compared to radiolabelling data during heat shock, as published before (170). Altogether this suggests that the effect of azhal-labelling on protein stability, on the time-scale employed for pulse-labelling, is negligible.

We also assume that the average k_{deg} is similar under aerobic and anaerobic conditions in our experiment. Evidence that this is a reasonable assumption comes from the quantitation of only methionine containing peptides. As mentioned above these peptides

represent only pre-existing material made before the pulse-labelling period. A quantitation of the protein levels derived from methionine peptide ratios is a measure for the relative stability of these pre-existing protein during the pulse-labelling time under anaerobic versus aerobic conditions. In supplemental Figure 3 the relative ratio of pre-existing protein material is presented, and it is evident that no large stability differences occur during pulse-labelling as the ratios do not differ significantly from 1 (0 on ^{2}log scale) for the vast majority of proteins.

Finally we assume that biomass is equal to total protein or $F(t) = P_{total}(t)$ in equation (a) and as $P_{total}(t) = P_{old} + P_{new}$ and the degradation of pre-existing proteins does not influence new protein formation, the rate of change of azhal-containing proteins P_{new} can be defined by differential equation (c).

$$\frac{\partial P_{new}}{\partial t} = k_{syn}P_{old} + (k_{syn} - k_{\deg})P_{new} \qquad (c)$$

Integration of the set of equations (b) and (c) is straightforward and yields the time-dependant functions $P_{old}(t)$ and $P_{new}(t)$. Boundary value conditions are at t = 0: $P_{old}(t) = P_{old}(0)$ and $P_{new}(t) = P_{new}(0) = 0$.

$$P_{old}(t) = P_{old}(0)e^{-k_{\deg}t} \qquad (d)$$

$$P_{new}(t) = P_{old}(0)[e^{(k_{syn}-k_{\deg})t} - e^{-k_{\deg}t}] \qquad (e)$$

And summation of (d) and (e)

$$P_{total}(t) = P_{old}(t) + P_{new}(t) = P_{old}(0)e^{(k_{syn}-k_{\deg})t} \qquad (f)$$

In our experiments we obtain two ratios. The first is the ratio between the number of newly synthesized proteins under either aerobiosis or anaerobiosis during the pulse-labelling time interval going from one environmental condition to the other. This ratio is derived from the iTRAQ reporter ions of azhal- containing peptides that are collected off-diagonal in our COFRADIC setup.

The predicted ratio of newly synthesized proteins under anaerobiosis $P_{new,\ anaerobiosis}(t)$ and aerobiosis $P_{new,\ aerobiosis}(t)$ is

$$R_{newly\_synthesized} = \frac{P_{new,anaerobiosis}(t)}{P_{new,aerobiosis}(t)} = \frac{P_{old,anaerobiosis}(0)[e^{(k_{syn,anaerobiosis}-k_{\deg})t} - e^{-k_{\deg}t}]}{P_{old,aerobiosis}(0)e^{(k_{syn,aerobiosis}-k_{\deg,})t} - e^{-k_{\deg}t}]} \qquad (g)$$

Which reduces to

$$R_{newly\_synthesized} = \frac{P_{old,anaerobiosis}}{P_{old,aerobiosis}} \frac{e^{k_{syn,anaerobiosis}t} - 1}{e^{k_{syn,aerobiosis}t} - 1} \quad \text{(h)}$$

For small values of $k_{syn}t$ as is the case in our study we can approximate the exponential function e^x by the first terms 1 + x of its Taylor expansion and the predicted ratio of newly synthesized proteins $P_{new}(t)$ becomes

$$R_{newly\_synthesized} = \frac{P_{new,anaerobiosis}(t)}{P_{new,aerobiosis}(t)} = \frac{P_{old,anaerobiosis}(0)}{P_{old,aerobiosis}(0)} \frac{k_{syn,anaerobiosis}}{k_{syn,aerobiosis}} \quad \text{(i)}$$

For small values of $k_{syn}t$, the measured ratio of azhal-containing peptides is predicted by the ratio of synthesis-rate constants in the aerobic and anaerobic experiment. Comparison with the data directly provides insight in differences of protein synthesis or translation rates between experimental conditions.

The second experimental ratio that we extract from our data set is the copy number ratio of total protein level at the end of the labelling time between two environmental conditions. This ratio is derived from the peptide ratios of peptides that do not contain azhal or methionine. Azhal peptides are excluded as these represent exclusively newly synthesized material, while methionine containing peptides are excluded because these represent only pre-existing material. Newly formed peptides that contain azhal are found in off-diagonal fractions. Peptides that do not contain azhal or methionine are found in both on-diagonal and off-diagonal fractions; these peptides reflect the total protein level in the cell as they are made up of both pre-existing as well as new material. The non azhal/methionine containing peptide copy number equals the summation of $P_{old}(t)$ and $P_{new}(t)$ in our kinetic model.

The predicted ratio R_{total_level} for non azhal/methionine containing peptides is

$$R_{total\_level} = \frac{P_{old,anaerobiosis}(t) + P_{new,anaerobiosis}(t)}{P_{old,aerobiosis}(t) + P_{new,aerobiosis}(t)} = \frac{P_{old,anaerobiosis}(0)}{P_{old,aerobiosis}(0)} \frac{e^{(k_{syn,anaerobiosis} - k_{deg})t}}{e^{(k_{syn,aerobiosis} - k_{deg})t}} \quad \text{(j)}$$

and after Taylor polynomial expansion

$$R_{total\_level} = \frac{P_{old,anaerobiosis}(0)}{P_{old,aerobiosis}(0)} \frac{[1 + (k_{syn,anaerobiosis} - k_{deg})t]}{[1 + (k_{syn,aerobiosis} - k_{deg})t]} \quad \text{(k)}$$

This total protein ratio reflects the overall protein expression ratio between two states. In our model we have two experimental ratios that are predicted by three reaction rate constants. To compare predicted ratios with measured ones we make the following assumptions. The aerobic culture is in the exponential growth phase. For an estimate of k_{deg} we assume a protein half-life of 120 minutes. From literature it is evident that with regard to degradation-rates we broadly find two protein populations. A small rapidly degrading

population with half-lives lower than our pulse time window and a more slowly degrading population with half-lives of at least a few hours up to more than 23 hours (110, 111). In our kinetic model we use the lower boundary of protein half-life $t_{½}$ = 120 minutes. For larger protein half-lives the effect of degradation on kinetics can be neglected. For $t_{½}$ = 120 minutes we calculate that k_{deg} = 5.78 x 10^{-3} min^{-1}. Next, we have to estimate $k_{syn.aerobisosis}$. During ten minutes labelling time the optical density (an approximation of biomass) of the culture increases with 6.25 %; assuming total protein content scales linearly with biomass its increase is also 6.25 %. Given our estimate of k_{deg} of 5.78 x 10^{-3} min^{-1} and a biomass increment of 6.25% we use equation (e) to calculate an average protein synthesis-rate constant $k_{syn,aerobic}$ = 1.2 x 10^{-2} min^{-1} under aerobic conditions.

For a few selected proteins with increased rates of new formation we used our model to estimate their $k_{syn,anaerobioses}$ rate constants using MATLAB (the Mathworks, Natick, USA). As an objective criterion we minimize the summation of the absolute value of the differences between the calculated ratios $R_{newly\_synthesized}$ and the measured one upon variation of $k_{syn,anaerobioses}$ To correct for differences in growth following the switch from aerobiosis to anaerobiosis we set $P_{old}(0)_{anaerobiosis}$ to 106.25 and $P_{old}(0)_{aerobiosis}$ to 100. In our method we mix absolute protein amounts of the azhal labelled anaerobic and aerobic proteomes in a one to one ratio.

The results are presented in Supplemental Table I. As is seen in this table our kinetic model predicts both protein level ratios reasonably well for the given set of first-order protein synthesis and degradation-rate constants and the estimated rate constants $k_{syn,anaerobiosis}$.

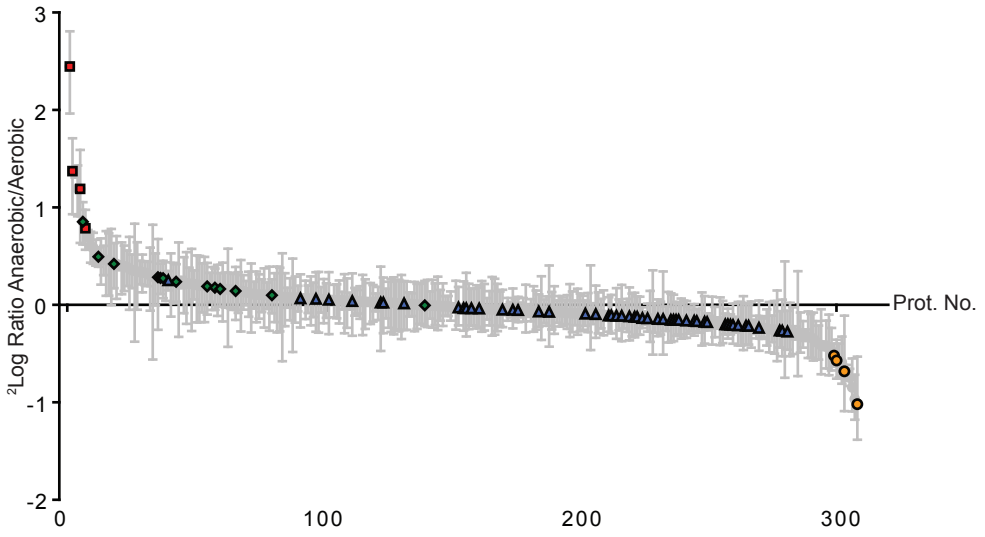
SUPPLEMENTAL TABLE I

Predicted and measured rates of formation of selected proteins and their change in levels after a switch to an anaerobic environment in E. coli.

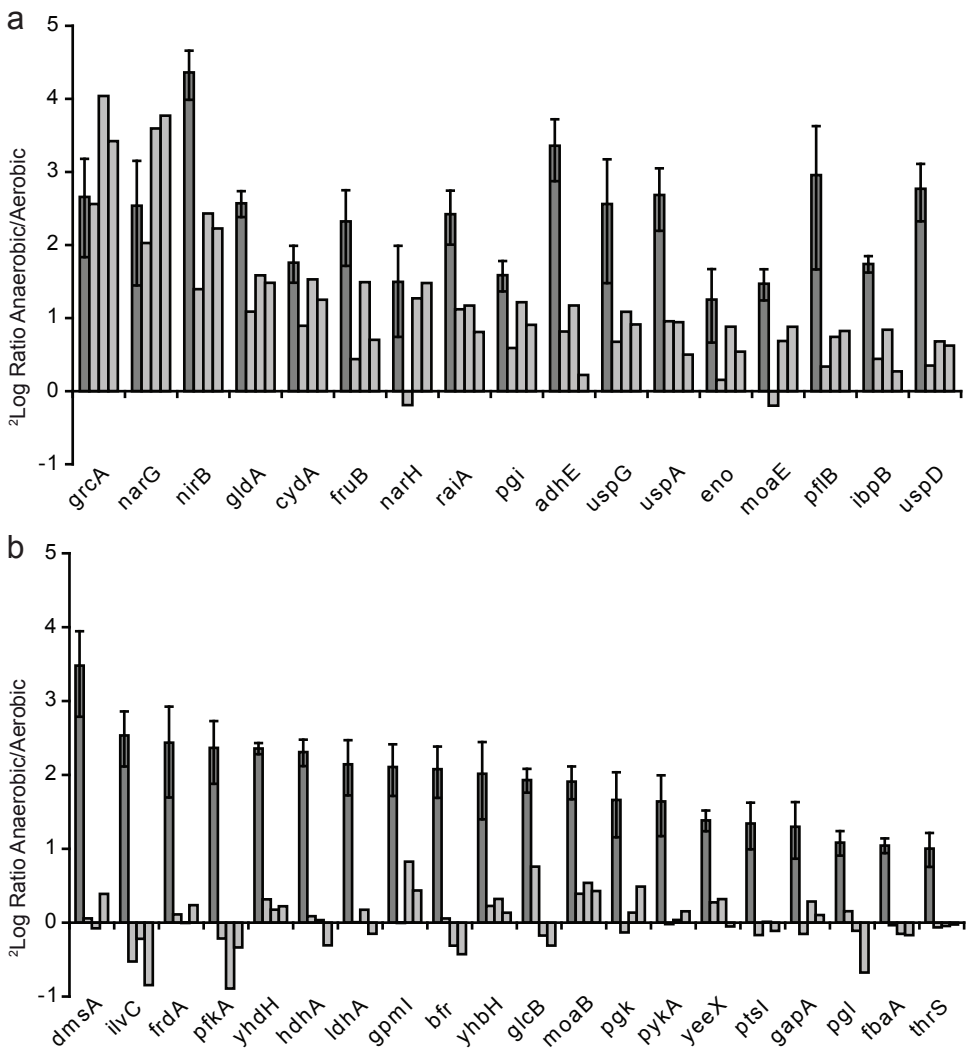
gene	|Δratio† |	|Δratio‡|	$k_{syn,\ anaerobiosis}$ (min^{-1})	ratio $k_{syn,\ anaerobiosis}$ / $k_{syn,\ aerobiosis}$
PfkA	0.08	0.09	4.7 x 10^{-2}	3.9
Gpml	0.12	0.01	4.1 x 10^{-2}	3.4
Pgk	0.04	0.02	3.3 x 10^{-2}	2.8
PykA	0.09	0.09	3.1 x 10^{-2}	2.6
Pgi	0.01	0.06	3.0 x 10^{-2}	2.5
GapA	0.06	0.03	2.5 x 10^{-2}	2.1
Eno	0.09	0.06	2.4 x 10^{-2}	2.0
FbaA	0.09	0.04	2.1 x 10^{-2}	1.8

|Δratio†| is the absolute value of the difference between the measured and predicted ratio anaerobic/aerobic of newly formed proteins determined by quantitation of azhal-tagged proteins, |Δratio‡| is the absolute value of the difference between the measured and predicted ratio anaerobic/aerobic of protein levels determined by quantitation of non-tagged azhal peptides.

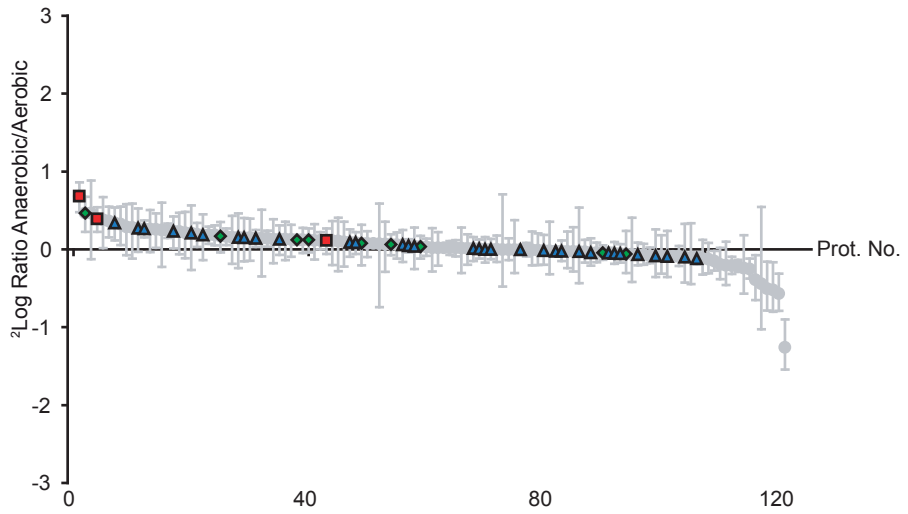
These rate constants are (as stated above) relative values that depend on initial conditions for $P_{old}(0)$, $k_{syn,\ aerobiosis}$ and k_{deg} for both growth conditions, where $k_{syn,aerobiosis}$ is determined by the growth rate of the *E. coli* population. Although our kinetic equations provide relative rate constants for protein synthesis, the ratio $k_{syn,anaerobises}/k_{syn,aerobiosis}$ is a direct measure for regulation of protein expression by adjusting the translation rate of proteins. The calculations also show that given our premises upon transition from aerobic to anaerobic growth conditions the ratio $k_{syn,anaerobises}/k_{syn,aerobiosis}$ increases by a factor of 1.8 to 3.9 and that the rate constant for degradation (k_{deg}) is an order of magnitude lower than the rate constants for protein synthesis (k_{syn}). Consequently, increased or decreased levels of most newly synthesized proteins result from changes in translation rate.



Supplemental Fig.1. **Changes in protein levels 10 minutes after the onset of anaerobiosis.** Grey dots show ²log ratio of protein levels from most up-regulated to most down-regulated under anaerobic conditions. Error bars show the ²log transformed standard deviation of the peptide ratios per protein. Red squares, levels of proteins involved in anaerobic respiration and mixed acid fermentation (adhE, pflB, grcA cydA), green diamonds, glycolytic enzymes and proteins of the PTS system (pfkA, gpmI, pgk, pykA, pgi, tpiA, gapA, eno, pykF, fbaA, gpmA, fruB and ptsI). Levels of the ribosomal proteins rpmG, rplJ, rplP, rplL, rpmD, rpmC, rpsN, rpsA, rpsO, rpsK, rplN, rpsG, rpsB, rpmA, rpsM, rplY, rplM, rpmB, rplV, rplK, rplF, rpsI, rplA, rpsP, rpsF, rplE, rplI, rpsJ, rpmF, rpsC, rpsH, rpsD, rpsE, rplR, rplO, rplW, rpsS, rplU, rpsR, rpmH, rplD, rplS, rplB, rpsQ, rpsL, rpsU, rplX, rplC, rpsT, rplQ and rplT (blue triangles). Orange dots, levels of some severe down-regulated proteins (sodA, katG, ahpC, ahpF and ompE). Listed proteins are depicted from left to right



Supplemental Fig. 2. **Relative synthesis-rates of up-regulated proteins compared with relative mRNA levels.** Proteins with a more than twofold increase of their relative synthesis-rate were compared to their mRNA levels as determined by Partridge *et al.* (267) upon micro-aerobic switch. Dark grey bars: relative ²log transformed ratio of newly synthesized proteins (as found by azhal-labeling of anaerobic over aerobically grown cultures) with their standard deviations (error bars). Light grey bars: relative transcript levels (²log-scale) at 5, 10 and 15 minutes (from left to right) in a low oxygen environment. Panel (**a**): proteins of which the sum of corresponding relative mRNAs levels at 5, 10 and 15 minutes after the onset of anaerobiosis is at least 1.4 (²log-scale). Panel (**b**): the sum of relative mRNA levels at 5, 10 and 15 minutes is 1.3 (²log-scale) or less.



Supplemental Fig. 3. **Relative stability of pre-existing protein population under anaerobic or aerobic conditions during the pulse labling period.** Grey dots show ^{2}log ratio of relative levels of protein obtained from only methionine containing peptides ordered from highest to lowest relative ratio. Error bars show the ^{2}log transformed standard deviation of the methionine-peptide ratios per protein. Methionine containing peptides represents protein species made before pulse-labeling commenced. The relative protein level derived from only methionine peptides ratios is an indication of the relative pre-existing protein stability under anaerobic versus aerobic conditions during the pulse-labeling. As is clear from the figure, the relative stability is the same under aerobic and anaerobic conditions as relative ratios do not differ significantly from 1 (0 on ^{2}log scale) for the vast majority of proteins. Red squares, levels of proteins involved in anaerobic respiration and mixed acid fermentation, green diamonds, glycolytic enzymes and proteins of the PTS system Levels of the ribosomal proteins are shown by blue triangles.

Ab.

List of Abbreviations

ARCO	azide reactive cyclo-octyne
ATP	adenine tri-phosphate
CID	collision induced dissociation
CNBr	cyanogen bromide
COFRADIC	combined fractional diagonal chromatography
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
ESI	electrospray ionisation
FTMS	Fourier transform mass spectrometry
GO	gene ontology
HEK	human endothelial kidney
HPG	homopropargylglycine
IPTG	isopropyl β-D-1-thiogalactopyranoside
iTRAQ	isobaric tagging reagents for relative and absolute quantitation
KDG	2-keto-3-deoxy-gluconate
LB	Lysogeny broth
LC	liquid chromatography
M	molar
MALDI	matrix assisted laser desorption ionisation
2ME	2-mercapto-ethanol
MS	mass spectrometry
NAD^+	nicotinamide adenine di-nucleotide, oxidized form
NADH	nicotinamide adenine di-nucleotide, reduced form
OD_{600}	optical density, 600 nm
PDH	pyruvate dehydrogenase
PFL	pyruvate formate lyase
PTS	phosphotransferase system
Q	quadrupole
RNA	ribonucleic acid

miRNA	micro ribonucleic acid
mRNA	messenger ribonucleic acid
tRNA	transfer ribonucleic acid
rRNA	ribosomal ribonucleic acid
sRNA	small regulatory ribonucleic acid
RP-HPLC	reversed phase high performance liquid chromatography
SAM	*s*-adenosyl methionine
SORI	sustained off-resonance irradiation
TCEP	tris(2-carboxy-ethyl)-phosphine
TFA	trifluoroacetic acid
TOF	time of flight
UV-VIS	ultra violet-visible light

R

References

1. E. S. Lander *et al.*, *Nature* **409**, 860 (2001).
2. J. C. Venter *et al.*, *Science* **291**, 1304 (2001).
3. P. Bertone *et al.*, *Science* **306**, 2242 (2004).
4. L. David *et al.*, *Proc Natl Acad Sci U.S.A.* **103**, 5320 (2006).
5. D. A. Lashkari *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13057 (1997).
6. M. Schena *et al.*, *Science* **270**, 467 (1995).
7. B. Alberts *et al.*, *Molecular Biology of the Cell*. (Garland Science, New York, ed. 5th, 2007), pp. 1,392.
8. L. Stryer *et al.*, *Biochemistry*. (W.H. Freeman, New York, ed. 5th, 2002), pp. 1100.
9. B. Futcher *et al.*, *Mol. Cell. Biol.* **19**, 7357 (1999).
10. S. Ghaemmaghami *et al.*, *Nature* **425**, 737 (2003).
11. D. Greenbaum *et al.*, *Genome Biol.* **4**, 117 (2003).
12. T. J. Griffin *et al.*, *Mol. Cell. Proteomics* **1**, 323 (2002).
13. S. P. Gygi *et al.*, *Mol. Cell. Biol.* **19**, 1720 (1999).
14. P. S. Lee *et al.*, *Biotechnol. Bioeng.* **84**, 834 (2003).
15. L. Nie *et al.*, *Biochem. Biophys. Res. Commun.* **339**, 603 (2006).
16. M. W. Schmidt *et al.*, *Mol. Syst. Biol.* **3**, 79 (2007).
17. Q. Tian *et al.*, *Mol. Cell. Proteomics* **3**, 960 (2004).
18. M. P. Washburn *et al.*, *Proc. Natl. Acad. Sci. U S A* **100**, 3107 (2003).
19. G. Chen *et al.*, *Mol. Cell. Proteomics.* **1**, 304 (2002).
20. T. Ideker *et al.*, *Science* **292**, 929 (2001).
21. Y. Taniguchi *et al.*, *Science* **329**, 533 (2010).
22. R. Aebersold *et al.*, *Nature* **422**, 198 (2003).
23. M. Bantscheff *et al.*, *Anal. Bioanal. Chem.* **389**, 1017 (2007).
24. S. P. Gygi *et al.*, *Curr. Opin. Chem. Biol.* **4**, 489 (2000).
25. S. E. Ong *et al.*, *Nat. Chem. Biol.* **1**, 252 (2005).
26. T. Ideker *et al.*, *Annu. Rev. Genomics Hum. Genet.* **2**, 343 (2001).
27. H. Kitano, *Science* **295**, 1662 (2002).
28. T. Maier *et al.*, *FEBS Lett.* **583**, 3966 (2009).
29. P. S. Hegde *et al.*, *Curr. Opin. Biotechnol.* **14**, 647 (2003).
30. J. C. Marioni *et al.*, *Genome Res.* **18**, 1509 (2008).
31. A. Mortazavi *et al.*, *Nat. Methods* **5**, 621 (2008).
32. U. Nagalakshmi *et al.*, *Science* **320**, 1344 (2008).
33. Z. Wang *et al.*, *Nat. Rev. Genet.* **10**, 57 (2009).
34. F. Gebauer *et al.*, *Nat. Rev. Mol. Cell. Biol.* **5**, 827 (2004).
35. R. E. Halbeisen *et al.*, *Cell. Mol. Life Sci.* **65**, 798 (2008).
36. M. Kozak, *Gene* **234**, 187 (1999).
37. M. Kozak, *Gene* **361**, 13 (2005).
38. P. Sunnerhagen, *Mol. Genet. Genomics* **277**, 341 (2007).
39. J. W. Hershey, *Annu. Rev. Biochem.* **60**, 717 (1991).
40. T. E. Dever *et al.*, *Cell* **68**, 585 (1992).
41. K. Moldave, *Annu. Rev. Biochem.* **54**, 1109 (1985).
42. J. Shine *et al.*, *Nature* **254**, 34 (1975).
43. W. F. Jacob *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4757 (1987).
44. G. Weyens *et al.*, *J. Mol. Biol.* **204**, 1045 (1988).
45. B. Schauder *et al.*, *Gene* **78**, 59 (1989).
46. M. Kozak, *Nucleic Acids Res.* **15**, 8125 (1987).
47. H. Du *et al.*, *J. Biol. Chem.* **273**, 20494 (1998).
48. T. Yura *et al.*, *Annu. Rev. Microbiol.* **47**, 321 (1993).
49. H. Nagai *et al.*, *Biochimie* **73**, 1473 (1991).
50. H. Nagai *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10515 (1991).
51. L. Jenner *et al.*, *Science* **308**, 120 (2005).
52. A. M. Fallon *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3411 (1979).

53. M. Nomura *et al.*, *Annu. Rev. Biochem.* **53**, 75 (1984).
54. J. M. Zengel *et al.*, *Prog. Nucleic. Acid. Res. Mol Biol.* **47**, 331 (1994).
55. K. Gausing, *J. Mol. Biol.* **115**, 335 (1977).
56. S. Gottesman, *Annu. Rev. Microbiol.* **58**, 303 (2004).
57. L. S. Waters *et al.*, *Cell* **136**, 615 (2009).
58. W. Filipowicz *et al.*, *Nat. Rev. Genet.* **9**, 102 (2008).
59. G. Storz *et al.*, *Annu. Rev. Biochem.* **74**, 199 (2005).
60. G. Storz *et al.*, *Curr. Opin. Microbiol* **7**, 140 (2004).
61. S. Gottesman, *Trends Genet.* **21**, 399 (2005).
62. K. Papenfort *et al.*, *Res. Microbiol.* **160**, 278 (2009).
63. M. Kozak, *Mol. Cell. Biol.* **9**, 5134 (1989).
64. S. L. Wolin *et al.*, *EMBO J.* **7**, 3559 (1988).
65. J. Minshull *et al.*, *Nucleic Acids Res.* **14**, 6433 (1986).
66. S. Pedersen, *EMBO J.* **3**, 2895 (1984).
67. M. A. Sorensen *et al.*, *J. Mol. Biol.* **207**, 365 (1989).
68. M. Gouy *et al.*, *Nucleic Acids Res.* **10**, 7055 (1982).
69. R. Grantham *et al.*, *Nucleic Acids Res.* **8**, 1893 (1980).
70. C. G. Kurland, *FEBS Lett.* **285**, 165 (1991).
71. A. G. Ryazanov *et al.*, *Nature* **334**, 170 (1988).
72. A. G. Ryazanov *et al.*, *FEBS Lett.* **285**, 170 (1991).
73. R. H. Buckingham, *Biochimie* **76**, 351 (1994).
74. D. R. Cavener *et al.*, *Nucleic Acids Res.* **19**, 3185 (1991).
75. K. K. McCaughan *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 5431 (1995).
76. E. S. Poole *et al.*, *EMBO J.* **14**, 151 (1995).
77. W. P. Tate *et al.*, *Biochem. Cell. Biol.* **73**, 1095 (1995).
78. R. E. Halbeisen *et al.*, *PLoS Biol.* **7**, e105 (2009).
79. T. Inada *et al.*, *RNA* **8**, 948 (2002).
80. V. L. MacKay *et al.*, *Mol. Cell. Proteomics* **3**, 478 (2004).
81. N. T. Ingolia *et al.*, *Science* **324**, 218 (2009).
82. B. Dahlmann *et al.*, *FEBS Lett.* **251**, 125 (1989).
83. T. Tamura *et al.*, *Curr. Biol.* **5**, 766 (1995).
84. R. De Mot *et al.*, *Trends Microbiol.* **7**, 88 (1999).
85. A. P. Arrigo *et al.*, *Nature* **331**, 192 (1988).
86. P. E. Falkenburg *et al.*, *Nature* **331**, 190 (1988).
87. D. Voges *et al.*, *Annu. Rev. Biochem.* **68**, 1015 (1999).
88. M. Bochtler *et al.*, *Annu. Rev. Biophys. Biomol. Struct.* **28**, 295 (1999).
89. L. Bedford *et al.*, *Trends Cell Biol.* **20**, 391 (2010).
90. S. Gottesman, *Annu. Rev. Genet.* **30**, 465 (1996).
91. A. M. Cuervo *et al.*, *J. Mol. Med.* **76**, 6 (1998).
92. J. F. Dice, *Autophagy* **3**, 295 (2007).
93. A. Bachmair *et al.*, *Science* **234**, 179 (1986).
94. D. K. Gonda *et al.*, *J. Biol. Chem.* **264**, 16700 (1989).
95. T. Potuschak *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7904 (1998).
96. J. W. Tobias *et al.*, *Science* **254**, 1374 (1991).
97. A. Mogk *et al.*, *Trends Cell. Biol.* **17**, 165 (2007).
98. A. Varshavsky, *Nat. Struct. Mol. Biol.* **15**, 1238 (2008).
99. A. Erbse *et al.*, *Nature* **439**, 753 (2006).
100. R. L. Ninnis *et al.*, *EMBO J.* **28**, 1732 (2009).
101. C. Giglione *et al.*, *Cell. Mol. Life Sci.* **61**, 1455 (2004).
102. S. Rogers *et al.*, *Science* **234**, 364 (1986).
103. F. Arsene *et al.*, *Int. J. Food Microbiol.* **55**, 3 (2000).

104. B. Bukau, *Mol. Microbiol.* **9**, 671 (1993).
105. E. Guisbert *et al.*, *Microbiol. Mol. Biol. Rev.* **72**, 545 (2008).
106. A. D. Grossman *et al.*, *Genes Dev.* **1**, 179 (1987).
107. D. B. Straus *et al.*, *Nature* **329**, 348 (1987).
108. H. Bahl *et al.*, *Genes Dev.* **1**, 57 (1987).
109. A. L. Goldberg *et al.*, *Annu. Rev. Biochem.* **45**, 747 (1976).
110. K. L. Larrabee *et al.*, *J. Biol. Chem.* **255**, 4125 (1980).
111. R. D. Mosteller *et al.*, *J. Biol. Chem.* **255**, 2524 (1980).
112. J. S. Andersen *et al.*, *Nature* **433**, 77 (2005).
113. R. J. Bateman *et al.*, *J. Am. Soc. Mass. Spectrom.* **18**, 997 (2007).
114. F. Bouwman *et al.*, *Proteomics* **4**, 3855 (2004).
115. B. J. Cargile *et al.*, *Anal. Chem.* **76**, 86 (2004).
116. M. K. Doherty *et al.*, *J. Proteome Res.* **8**, 104 (2009).
117. M. K. Doherty *et al.*, *Proteomics* **5**, 522 (2005).
118. N. Gustavsson *et al.*, *Proteomics* **5**, 3563 (2005).
119. J. M. Pratt *et al.*, *Mol. Cell. Proteomics* **1**, 579 (2002).
120. P. K. Rao *et al.*, *Anal. Chem.* **80**, 6860 (2008).
121. P. K. Rao *et al.*, *Anal. Chem.* **80**, 396 (2008).
122. P. G. Lemaux *et al.*, *Cell* **13**, 427 (1978).
123. S. Pedersen *et al.*, *Cell* **14**, 179 (1978).
124. M. W. Smith *et al.*, *J. Bacteriol.* **154**, 336 (1983).
125. C. Gerner *et al.*, *Mol. Cell. Proteomics* **1**, 528 (2002).
126. D. Hoper *et al.*, *Proteomics* **6**, 1550 (2006).
127. R. Dover *et al.*, *J. Cell Sci.* **107 (Pt 5)**, 1181 (1994).
128. V. W. Hu *et al.*, *FASEB J.* **14**, 448 (2000).
129. V. W. Hu *et al.*, *FASEB J.* **15**, 1562 (2001).
130. N. F. Marko *et al.*, *FASEB J.* **17**, 1470 (2003).
131. M. Yanokura *et al.*, *Int. J. Radiat. Biol.* **76**, 295 (2000).
132. J. Yeargin *et al.*, *Curr. Biol.* **5**, 423 (1995).
133. S. Zhou *et al.*, *Proteomics* **5**, 2739 (2005).
134. X. Zuo *et al.*, *Electrophoresis* **21**, 3035 (2000).
135. M. K. Doherty *et al.*, *Expert Rev. Proteomics* **3**, 97 (2006).
136. Q. Li, *Mass Spectrom. Rev.* **29**, 717 (2009).
137. J. A. Vogt *et al.*, *Anal. Chem.* **77**, 2034 (2005).
138. M. K. Hellerstein *et al.*, *Am. J. Physiol.* **263**, E988 (1992).
139. M. K. Hellerstein *et al.*, *Am. J. Physiol.* **276**, E1146 (1999).
140. C. Papageorgopoulos *et al.*, *Anal. Biochem.* **267**, 1 (1999).
141. G. Kramer *et al.*, *Mol. Cell. Proteomics* **8**, 1599 (2009).
142. T. L. Hendrickson *et al.*, *Annu. Rev. Biochem.* **73**, 147 (2004).
143. J. W. Back *et al.*, *Angew. Chem. Int. Ed. Engl.* **44**, 7946 (2005).
144. K. L. Kiick *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 19 (2002).
145. K. L. Kiick *et al.*, *Tetrahedron* **56**, 9487 (2000).
146. K. L. Kiick *et al.*, *Angew. Chem. Int. Ed.* **39**, 2148 (2000).
147. K. L. Kiick *et al.*, *FEBS Lett.* **502**, 25 (2001).
148. A. J. Link *et al.*, *J Am Chem Soc* **125**, 11164 (2003).
149. A. J. Link *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10180 (2006).
150. A. J. Link *et al.*, *J Am Chem Soc* **126**, 10598 (2004).
151. E. Strable *et al.*, *Bioconjugate Chem.* **19**, 866 (2008).
152. V. V. Rostovtsev *et al.*, *Angew. Chem. Int. Ed. Engl.* **41**, 2596 (2002).
153. C. W. Tornoe *et al.*, *J. Org. Chem.* **67**, 3057 (2002).
154. N. J. Agard *et al.*, *ACS Chem. Biol.* **1**, 644 (2006).

155. N. J. Agard *et al.*, *J Am Chem Soc* **126**, 15046 (2004).
156. J. A. Codelli *et al.*, *J Am Chem Soc* **130**, 11486 (2008).
157. E. M. Sletten *et al.*, *Org. Lett.* **10**, 3097 (2008).
158. H. Staudinger and J. Meyer, *Helv. Chim. Acta* **2**, 635 (1919).
159. E. Saxon *et al.*, *Org. Lett.* **2**, 2141 (2000).
160. E. Saxon *et al.*, *Science* **287**, 2007 (2000).
161. M. Kohn *et al.*, *Angew. Chem. Int. Ed. Engl.* **43**, 3106 (2004).
162. K. E. Beatty *et al.*, *Angew. Chem. Int. Ed. Engl.* **45**, 7364 (2006).
163. K. E. Beatty *et al.*, *Bioorg. Med. Chem. Lett.* **18**, 5995 (2008).
164. K. E. Beatty *et al.*, *J Am Chem Soc* **127**, 14150 (2005).
165. J. M. Baskin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 16793 (2007).
166. D. C. Dieterich *et al.*, *Nat. Neurosci* **13**, 897 (2010).
167. D. C. Dieterich *et al.*, *Nat. Protoc.* **2**, 532 (2007).
168. D. C. Dieterich *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9482 (2006).
169. G. Kramer *et al.*, *Mol. Cell. Proteomics* **9**, 2508 (2010).
170. M. A. Nessen *et al.*, *J. Proteome Res.* **8**, 3702 (2009).
171. R. B. Deal *et al.*, *Science* **328**, 1161 (2010).
172. G. Kramer, Master Thesis, University of Amsterdam (2004).
173. M. A. Nessen, PhD-Thesis, University of Amsterdam (2010).
174. J. B. Mangold *et al.*, *Mutat. Res.* **216**, 27 (1989).
175. A. Wang *et al.*, *Chembiochem* **9**, 324 (2008).
176. S. Schoffelen *et al.*, *Bioconjug. Chem.* **19**, 1127 (2008).
177. T. E. Meyer, *Biochim. Biophys. Acta* **806**, 175 (1985).
178. S. Masuda *et al.*, *Cell* **110**, 613 (2002).
179. A. Losi *et al.*, *Biophys. J.* **82**, 2627 (2002).
180. M. Thanbichler *et al.*, *J. Bacteriol.* **181**, 662 (1999).
181. S. E. Ong *et al.*, *Mol. Cell. Proteomics* **1**, 376 (2002).
182. J. W. Gouw *et al.*, *Mol. Cell. Proteomics* **9**, 11 (2010).
183. P. L. Ross *et al.*, *Mol. Cell. Proteomics* **3**, 1154 (2004).
184. A. Schmidt *et al.*, *Proteomics* **5**, 4 (2005).
185. M. B. Smolka *et al.*, *Anal. Biochem.* **297**, 25 (2001).
186. A. Thompson *et al.*, *Anal. Chem.* **75**, 1895 (2003).
187. W. D. Hoff *et al.*, *Biophys. J.* **67**, 1691 (1994).
188. T. E. Meyer *et al.*, *Biophys. J.* **59**, 988 (1991).
189. T. E. Meyer *et al.*, *Biochemistry* **26**, 418 (1987).
190. S. Devanathan *et al.*, *Biochemistry* **37**, 11563 (1998).
191. M. Kumauchi *et al.*, *J. Biochem.* **132**, 205 (2002).
192. W. D. Hoff *et al.*, *Biochim. Biophys. Act. Bioenerg.* **1322**, 151 (1997).
193. M. A. van der Horst *et al.*, *FEBS Lett.* **497**, 26 (2001).
194. W. Laan *et al.*, *Biochemistry* **45**, 51 (2006).
195. A. Jung *et al.*, *J. Mol. Biol.* **362**, 717 (2006).
196. S. Masuda *et al.*, *Plant Cell Physiol.* **49**, 1600 (2008).
197. K. Sadeghian *et al.*, *J Am Chem Soc* **130**, 12501 (2008).
198. J. T. Lundquist *et al.*, *Org. Lett.* **3**, 781 (2001).
199. R. Kort *et al.*, *EMBO J.* **15**, 3209 (1996).
200. W. Laan *et al.*, *Photochem. Photobiol.* **78**, 290 (2003).
201. M. Avila-Perez *et al.*, *J. Bacteriol.* **188**, 6411 (2006).
202. J. H. Miller, *Experiments in Molecular Genetics*. (Cold Spring Harbor Laboratories, New York, 1972).
203. J. Spizizen, *Proc. Natl. Acad. Sci. U.S.A.* **44**, 1072 (1958).

204. J. Hendriks *et al.*, *Biophys. J.* **82**, 1632 (2002).
205. A. M. Brown, *Comput. Methods Programs Biomed.* **65**, 191 (2001).
206. B. P. Nichols *et al.*, *J. Bacteriol.* **180**, 6408 (1998).
207. M. Singer *et al.*, *Microbiol. Rev.* **53**, 1 (1989).
208. B. A. McDaniel *et al.*, *J. Bacteriol.* **188**, 3674 (2006).
209. A. Losi *et al.*, *Photochem. Photobiol. Sci.* **2**, 759 (2003).
210. Y. Tang *et al.*, *Photochem. Photobiol. Sci.* **9**, 47 (2010).
211. M. Gauden *et al.*, *Biochemistry* **44**, 3653 (2005).
212. B. J. Kraft *et al.*, *Biochemistry* **42**, 6726 (2003).
213. K. Gevaert *et al.*, *Mol. Cell. Proteomics* **1**, 896 (2002).
214. K. Gevaert *et al.*, *Proteomics* **4**, 897 (2004).
215. K. Gevaert *et al.*, *Nat. Biotechnol.* **21**, 566 (2003).
216. K. Gevaert *et al.*, *Proteomics* **5**, 3589 (2005).
217. B. Ghesquiere *et al.*, *J. Proteome Res.* **5**, 2438 (2006).
218. B. Ghesquiere *et al.*, *Rapid Commun. Mass Spectrom.* **20**, 2885 (2006).
219. X. Hanoulle *et al.*, *J. Prot. Res.* **5**, 3438 (2006).
220. I. L. Cartwright *et al.*, *Nucleic Acids Res.* **3**, 2331 (1976).
221. J. V. Staros *et al.*, *Biochem. Biophys. Res. Commun.* **80**, 568 (1978).
222. J. B. C. Findlay, Geisow, M. J. , *Protein Sequencing: A Practical Approach*. (IRL, Oxford, 1989).
223. P. T. Kasper *et al.*, *Chembiochem* **8**, 1281 (2007).
224. R. W. Corbin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9232 (2003).
225. J. L. Robbins-Manke *et al.*, *J. Bacteriol.* **187**, 7027 (2005).
226. A. Diaz-Acosta *et al.*, *Arch. Microbiol.* **185**, 429 (2006).
227. R. Landick *et al.*, *Cell* **38**, 175 (1984).
228. T. Yura *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **81**, 6803 (1984).
229. G. Nonaka *et al.*, *Genes Dev.* **20**, 1776 (2006).
230. K. Zhao *et al.*, *J. Biol. Chem.* **280**, 17758 (2005).
231. H. D. Meiring *et al.*, *J. Sep. Sci.* **25**, 557 (2002).
232. B. R. Zeeberg *et al.*, *Genome. Biol.* **4**, (2003).
233. S. Cooper *et al.*, *Mol. Gen. Genet.* **139**, 167 (1975).
234. S. W. Harcum *et al.*, *J. Ind. Microbiol. Biotechnol.* **33**, 801 (2006).
235. C. S. Richmond *et al.*, *Nucleic. Acids Res.* **27**, 3821 (1999).
236. J. L. Brissette *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 862 (1990).
237. J. L. Brissette *et al.*, *J. Mol. Biol.* **220**, 35 (1991).
238. A. J. Darwin, *Mol. Microbiol.* **57**, 621 (2005).
239. P. Model *et al.*, *Mol. Microbiol.* **24**, 255 (1997).
240. L. Weiner *et al.*, *Genes Dev.* **5**, 1912 (1991).
241. T. Yamamori *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **79**, 860 (1982).
242. S. A. Bissonnette *et al.*, *Mol. Microbiol.* **75**, 1539 (2010).
243. N. Peekhaus *et al.*, *J. Bacteriol.* **180**, 3495 (1998).
244. T. Hunt *et al.*, *Mol. Cell. Proteomics* **3**, S286 (2004).
245. A. M. Boehm *et al.*, *BMC Bioinformatics* **8**, 214 (2007).
246. F. E. Satterthwaite, *Inter. Biom. Soc.* **2**, 110 (1946).
247. Y. Benjamini *et al.*, *J. R. Stat. Soc. Ser. B Stat. Methodol.* **57**, 289 (1995).
248. L. Argaman *et al.*, *Curr. Biol.* **11**, 941 (2001).
249. S. Chen *et al.*, *Biosystems* **65**, 157 (2002).
250. E. Rivas *et al.*, *Curr. Biol.* **11**, 1369 (2001).
251. K. M. Wassarman *et al.*, *Gen. Dev.* **15**, 1637 (2001).
252. H. Aiba, *Curr. Opin. Microbiol.* **10**, 134 (2007).
253. G. Unden *et al.*, *Biochim. Biophys. Acta* **1320**, 217 (1997).

254. D. P. Clark, *FEMS Microb. Rev.* **63**, 223 (1989).
255. S. Alexeeva *et al.*, *J. Bacteriol.* **185**, 204 (2003).
256. C. Kisker *et al.*, *Annu. Rev. Biochem.* **66**, 233 (1997).
257. G. Sulzenbacher *et al.*, *Acta. Crystallogr. D. Biol. Crystallogr.* **60**, 1855 (2004).
258. P. A. Cotter *et al.*, *Mol. Microbiol.* **25**, 605 (1997).
259. A. F. V. Wagner *et al.*, *Biochem. Biophys. Res. Commun.* **285**, 456 (2001).
260. K. Kvint *et al.*, *Curr. Opin. Microbiol.* **6**, 140 (2003).
261. M. Ueta *et al.*, *Genes Cells* **10**, 1103 (2005).
262. W. S. Champney, *Mol. Gen. Genet.* **152**, 259 (1977).
263. R. J. Harvey, *J. Bacteriol.* **101**, 574 (1970).
264. S. Gottesman *et al.*, *J. Biol. Chem.* **265**, 7886 (1990).
265. Z. Maglica *et al.*, *J. Mol. Biol.* **384**, 503 (2008).
266. T. Gaal *et al.*, *Science* **278**, 2092 (1997).
267. J. D. Partridge *et al.*, *J. Biol. Chem.* **282**, 11230 (2007).
268. J. A. Bernstein *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2758 (2004).
269. Y. Wang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5860 (2002).
270. S. S. van Berkel *et al.*, *Chembiochem* **8**, 1504 (2007).
271. M. Gilar *et al.*, *J. Sep. Sci.* **28**, 1694 (2005).
272. J. Szychowski *et al.*, *J Am Chem Soc* **132**, 18351 (2010).
273. A. Sittka *et al.*, *PLoS Genet.* **4**, e1000163 (2008).
274. K. Kimata *et al.*, *EMBO J.* **20**, 3587 (2001).
275. C. K. Vanderpool *et al.*, *Mol. Microbiol.* **54**, 1076 (2004).
276. F. M. Ausubel *et al.*, *Current Protocols in Molecular Biology*, (Wiley Interscience, 1998).
277. L. Martens *et al.*, *Bioinformatics* **21**, 3584 (2005).
278. R. D. Mosteller *et al.*, *J. Theor. Biol.* **108**, 597 (1984).
279. A. L. Goldberg, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 422 (1972).
280. M. J. Pine, *Antimicrobial Agents and Chemotherapy* **13**, 676 (1978).

List of publications

Kramer, G., Kasper, P. T., de Jong, L., and de Koster, C. G. (2011) *Quantitation of newly synthesized proteins by pulse labeling with azidohomoalanine.* Methods Mol. Biol. 753, 169-181. (**c**)

Aerts, J. M., Kallemeijn, W. W., Wegdam, W., Joao Ferraz, M., van Breemen, M. J., Dekker, N., **Kramer, G.**, Poorthuis, B. J., Groener, J. E., Cox-Brinkman, J., Rombach, S. M., Hollak, C. E., Linthorst, G. E., Witte, M. D., Gold, H., van der Marel, G. A., Overkleeft, H. S., and Boot, R. G. (2011) *Biomarkers in the diagnosis of lysosomal storage disorders: proteins, lipids, and inhibodies.* J. Inherit. Metab. Dis. 34, 605-619.

Kramer, G., Sprenger, R. R., Nessen, M. A., Roseboom, W., Speijer, D., de Jong, L., de Mattos, M. J., Back, J., and de Koster, C. G. (2010) *Proteome-wide alterations in Escherichia coli translation rates upon anaerobiosis.* Mol. Cell. Proteomics 9, 2508-2516. (**d**)

Witte, M. D., Kallemeijn, W. W., Aten, J., Li, K. Y., Strijland, A., Donker-Koopman, W. E., van den Nieuwendijk, A. M., Bleijlevens, B., **Kramer, G.**, Florea, B. I., Hooibrink, B., Hollak, C. E., Ottenhoff, R., Boot, R. G., van der Marel, G. A., Overkleeft, H. S., and Aerts, J. M. (2010) *Ultrasensitive in situ visualization of active glucocerebrosidase molecules.* Nat. Chem. Biol. 6, 907-913.

Kramer, G., Sprenger, R. R., Back, J., Dekker, H. L., Nessen, M. A., van Maarseveen, J. H., de Koning, L. J., Hellingwerf, K. J., de Jong, L., and de Koster, C. G. (2009) *Identification and quantitation of newly synthesized proteins in Escherichia coli by enrichment of azidohomoalanine-labeled peptides with diagonal chromatography.* Mol. Cell. Proteomics 8, 1599-1611. (**a**)

Nessen, M. A., **Kramer, G.**, Back, J., Baskin, J. M., Smeenk, L. E., de Koning, L. J., van Maarseveen, J. H., de Jong, L., Bertozzi, C. R., Hiemstra, H., and de Koster, C. G. (2009) *Selective enrichment of azide-containing peptides from complex mixtures*. J. Proteome Res. 8, 3702-3711.

Bekker, M., **Kramer, G.**, Hartog, A. F., Wagner, M. J., de Koster, C. G., Hellingwerf, K. J., and de Mattos, M. J. (2007) *Changes in the redox state and composition of the quinone pool of Escherichia coli during aerobic batch-culture growth.* Microbiology 153, 1974-1980.

Back, J. W., David, O., **Kramer, G.**, Masson, G., Kasper, P. T., de Koning, L. J., de Jong, L., van Maarseveen, J. H., and de Koster, C. G. (2005) *Mild and chemoselective peptide-bond cleavage of peptides and proteins at azido homoalanine.* Angew. Chem. Int. Ed. Engl. 44, 7946-7950. (**b**)

Cremazy, F. G., Manders, E. M., Bastiaens, P. I., **Kramer, G.**, Hager, G. L., van Munster, E. B., Verschure, P. J., Gadella, T. J., Jr., and van Driel, R. (2005) *Imaging in situ protein-DNA interactions in the cell nucleus using FRET-FLIM*. Exp. Cell. Res. 309, 390-396.

de Ruijter, A. J., Kemp, S., **Kramer, G**., Meinsma, R. J., Kaufmann, J. O., Caron, H. N., and van Kuilenburg, A. B. (2004) *The novel histone deacetylase inhibitor BL1521 inhibits proliferation and induces apoptosis in neuroblastoma cells.* Biochem. Pharmacol. 68, 1279-1288.

Chapters

Dankwoord

Wanneer je `dankwoord proefschrift' zoekt in Google vindt je ongeveer 55.400 resultaten. Naast een artikel uit 2010 in het Leids universitair weekblad waarin gewag wordt gemaakt dat het bij de universiteit Leiden niet langer is toegestaan 'op niet terughoudende wijze God, huisdieren of de hardloop club te danken', blijken vele daadwerkelijke dankwoorden van proefschriften te zijn. In zoverre je een zo groot aantal dankwoorden zou kunnen lezen blijken in ieder geval een groot aantal resultaten op de eerste vijf pagina's (gelukkig dat Google relevantie voor ons kan bepalen) een aantal mensen te bedanken die geholpen hebben bij het tot stand komen van het proefschrift. Woorden als 'teamwork' en 'bijdrage van velen' worden hierbij niet geschroomd. En zo is het ook, wetenschap bedrijf je zelden alleen.

En zo dank ik ten eerste mijn promotor, voor de mogelijkheid mijn promotie op dit onderwerp te starten en de ruimte die je altijd gegeven hebt voor het najagen van eigen ideeën evenals de grote hulp die je bent geweest bij het modelleren van eiwitsynthese en het oplossen van massaspectra van kleine moleculen die in samenwerkingsverbanden gemeten dienden te worden.

Daarnaast zijn mijn beide copromotores cruciaal geweest in verscheidene opzichten, zoals de inval waarmee het project van start ging als ook duizend en één goede, gemiddelde of minder goede ideeën hoe we azhal bevattende peptiden zouden opzuiveren. Of het blijven wijzen op de mogelijkheid van diagonale chromatografie als goede manier van azide peptide verrijking ondanks vele tegenwerpingen van deze promovendus, die het toch wel erg veel werk vond.

Uiteraard dank ik ook de commissie voor haar lezen en becommentariëren van dit werk, alsmede de oppositie die ik tegemoet kan zien, hetgeen voor sommigen een reis uit het buitenland vergt.

Dan kan ik natuurlijk niet de wetenschappers vergeten waarmee ik de afgelopen jaren heb samengewerkt, of het nu ging om het kweken van *B. subtilis* op azhal aan de Ohio State University of het meten van massa's van recombinante eiwitten, quinonen, riboflavines etc. van vooral toch de 8e verdieping, het maakt me tot de massaspectrometrist die ik vandaag de dag ben.

Uiteraard kan ik ook mijn collegae niet vergeten, bedankt voor de prettige werkomgeving die ik zowel op het 'Roeterseiland' als het AMC ervaar en heb ervaren. Dank dus, collega's van medische biochemie, microbiology (zowel normaal als voedselveiligheid) en natuurlijk massaspectrometrie. In het bijzonder de collega's met wie ik voor langere tijd het kantoor deel en heb gedeeld. En die collega's met wie ik elke dag weer mag lunchen. En niet te vergeten die collega's met wie ik te vaak te diep en tot te laat in het glas keek. Of de collega met wie ik altijd koffie drink. En natuurlijk de collega die me op een goedkope drukker

wees. Maar nu wordt het teveel een opsomming. Bedankt allemaal, ook als je niet tot de bovenstaande groepen behoort, je vulde vast een universiteitsgebouw met je aanwezigheid.

Ook wil ik mijn paranimfen in het bijzonder danken voor het feit dat ze me willen bijstaan tijdens deze laatste fase van de promotie. Ik hoop dat jullie niet daadwerkelijk in actie hoeven komen.

Beste familie en vrienden, dit is het dan. Ik hoop dat ik jullie niet teveel verveeld heb met verhalen over wat ik eigenlijk doe. Massaspectrometrie doet het niet altijd even goed op feestjes, uitzonderingen daargelaten (massaspectrometrie bijeenkomsten met een open bar).

Lieve moeder, dit is nu waar ik de laatste jaren mee bezig was. Dank voor alles, jij bent uiteraard het fundament van dit allemaal.

En natuurlijk mijn lieve echtgenote, Marcela. Dank voor het luisteren, het praten en de steun bij alles in mijn leven. With you I believe in some kind of path that we can walk down, me and you.

Rest mij als Amsterdams promovendus naar goed Leids gebruik het danken van God, huisdieren en de hardloop club achterwege te laten.

Dankwoord