Determination of protein synthesis on a proteomic scale

Kramer, G.

Publication date
2011

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

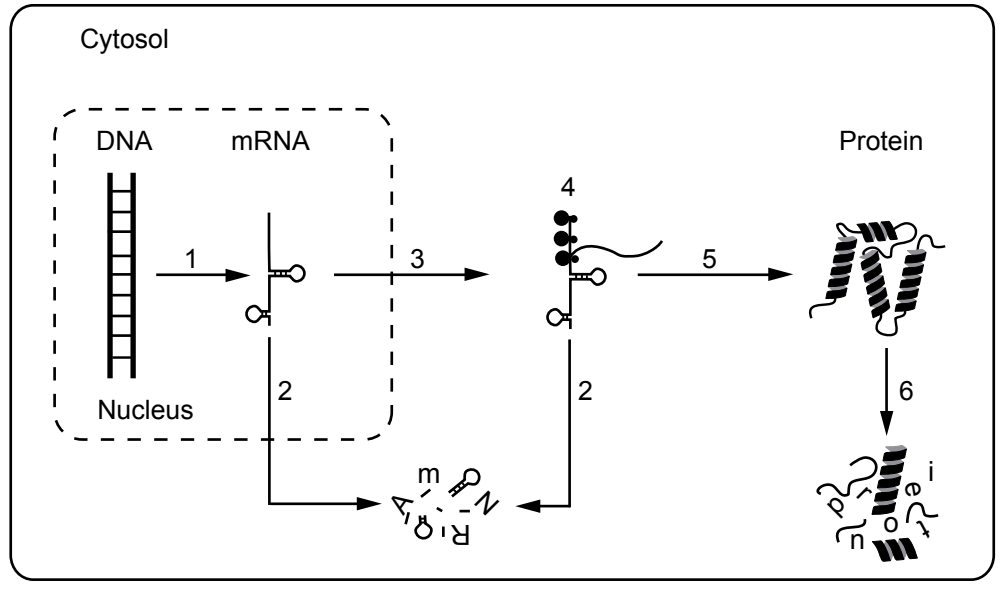

Figure 1. **From DNA to protein.** Gene expression starts with transcription (1) of genetic information, generating mRNA molecules, from which (in eukaryotes) introns are removed, a 5' cap and poly-A tail are added prior to export from the nucleus (3). mRNA levels measured in the cell are themselves 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

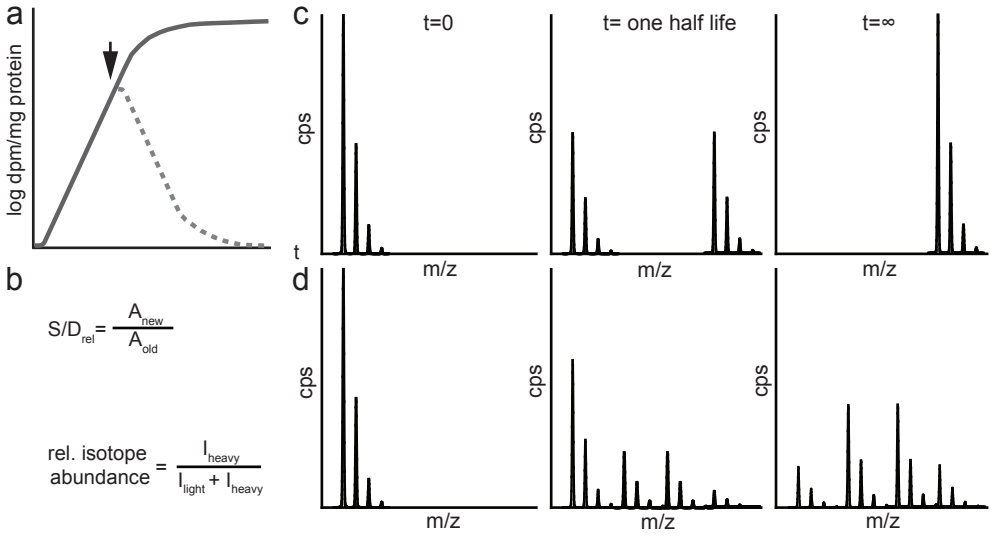

FIGURE 2. **Pulse-chase labelling measures protein synthesis and degradation.** Upon addition of a radiolabel at t=0 a short lag period follows as both unlabeled precursor needs to be depleted and radiolabel needs to be imported into the cell before it can be used in protein synthesis. Subsequently total 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

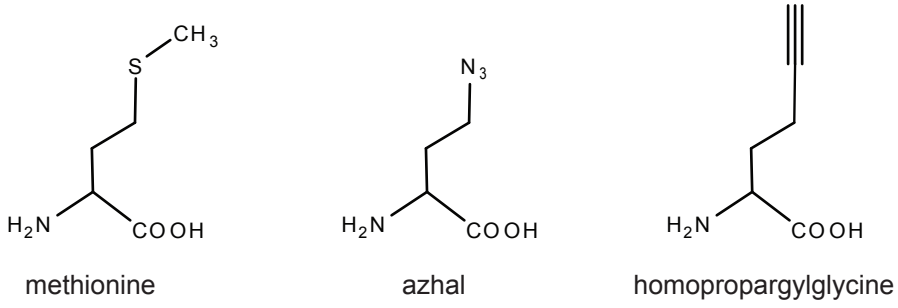

FIGURE 3. **Structure of methionine and two non-natural analogues.**

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%

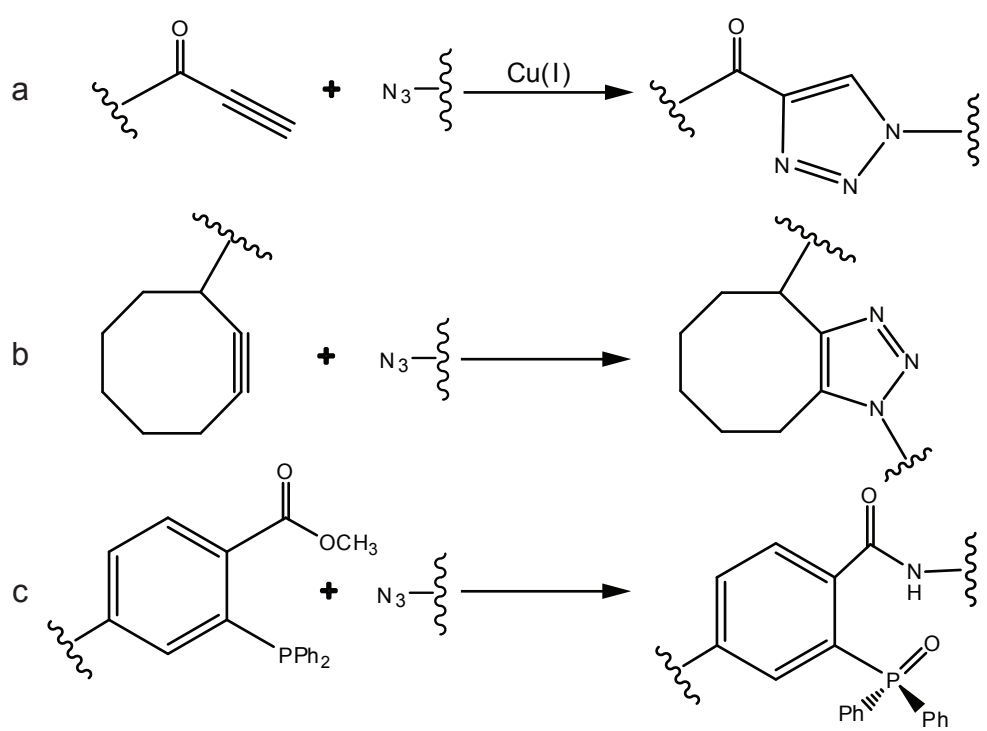

FIGURE 4. **Chemical reactions directed against azhal.** Copper catalyzed (3+2) cyclo-addition between terminal alkyne- and an azide-group (**a**). Strain-promoted (3+2) cyclo-addition between an octyne and an azide (**b**). Staudinger ligation involving an azide reacting with a phosphine (**c**).

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.

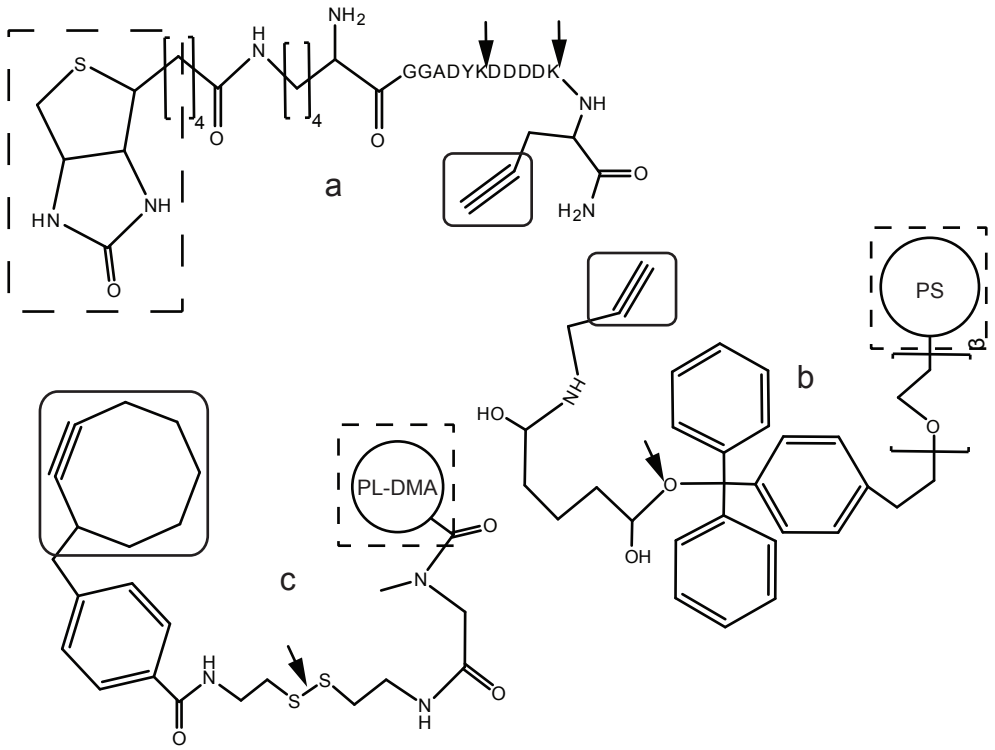

Figure 5. **Affinity resins for the selective enrichment of azhal containing molecules.** Terminal-alkyne biotin tag employed in the BONCAT approach described by Dieterich *et al.* (**a**). Terminal-alkyne coupled to a polystyrene bead (**b**). ARCO-resin employing an octyne reactive group coupled to 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.