Determination of protein synthesis on a proteomic scale

Kramer, G.

Publication date
2011

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

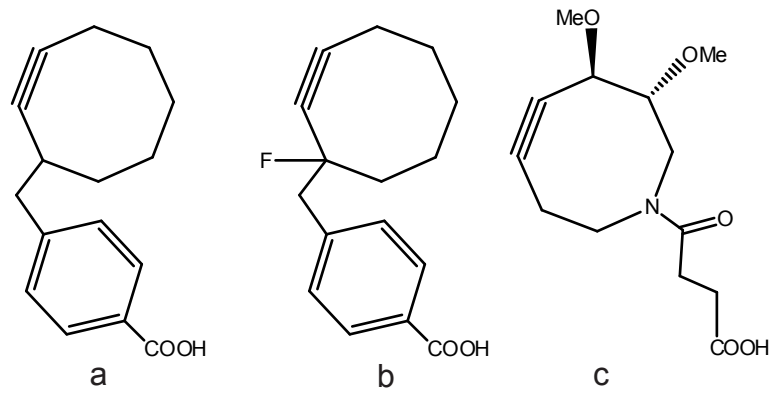

FIGURE 1. **Different cyclo-octynes used in azide-reactive affinity resins.** The cyclo-octyne (**a**) is successfully used to enrich azide-containing peptides with ARCO-resin as described by Nessen *et al.* (170), whereas use of the mono-fluorinated cyclo-octyne (**b**) or the more hydrophilic azacyclo-octyne (**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.