Determination of protein synthesis on a proteomic scale

Kramer, G.

Publication date
2011

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,

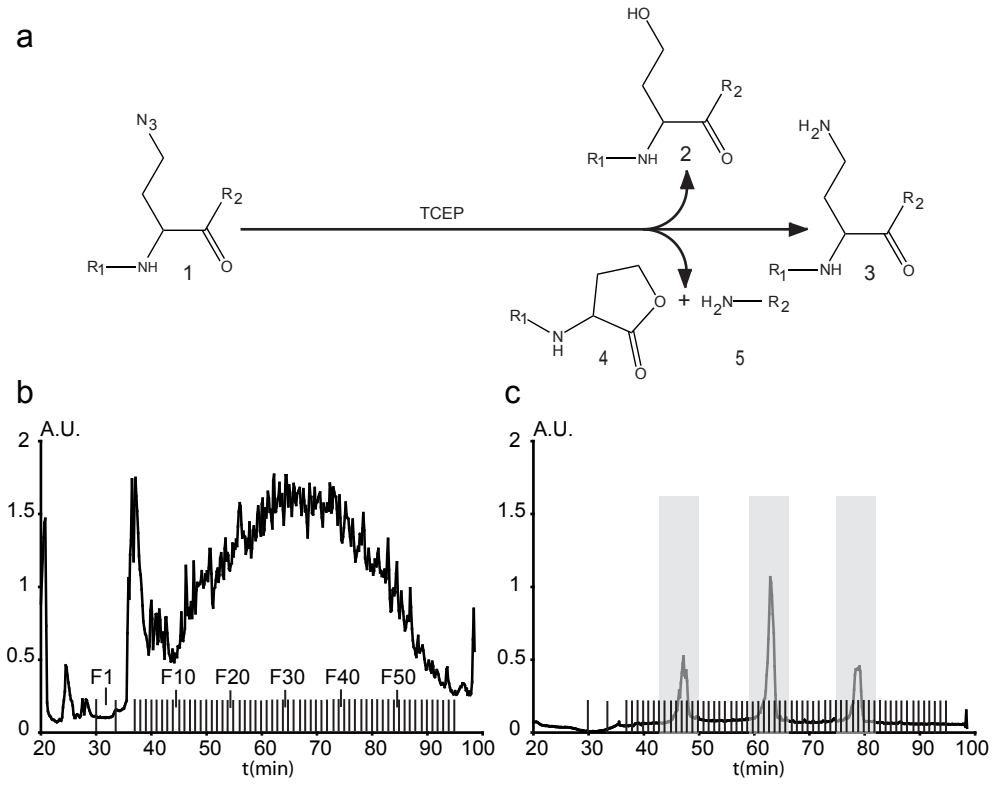

Figure 1. **The COFRADIC approach**. reactions products resulting from an azhal (**1**) -containing peptide reacting with TCEP are: a peptide containing either a homoserine residue (**2**), or a di-aminobutyrate residue (**3**), or an N-terminal cleavage product with a homoserine lactone residue at its 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.

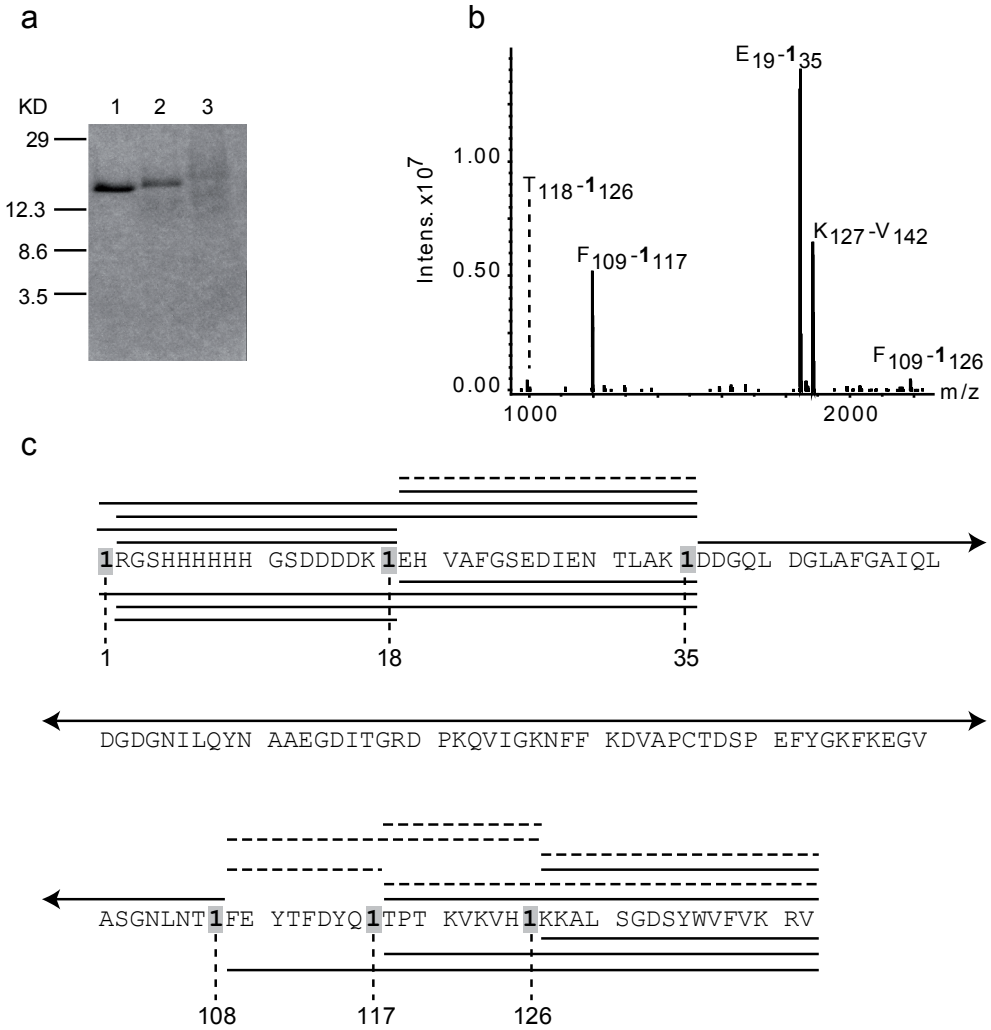

FIGURE 2. **Recombinant azPYP is cleaved by TCEP.** Coomassie Blue stained gel containing equal amounts (23 mg) of azPYP incubated without (lane 1) and with 10 mm (lane 2) and 100 mm (lane 3) TCEP at pH 5 (**a**); the intact protein disappears and is cut into smaller fragments that do not nicely resolve 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.

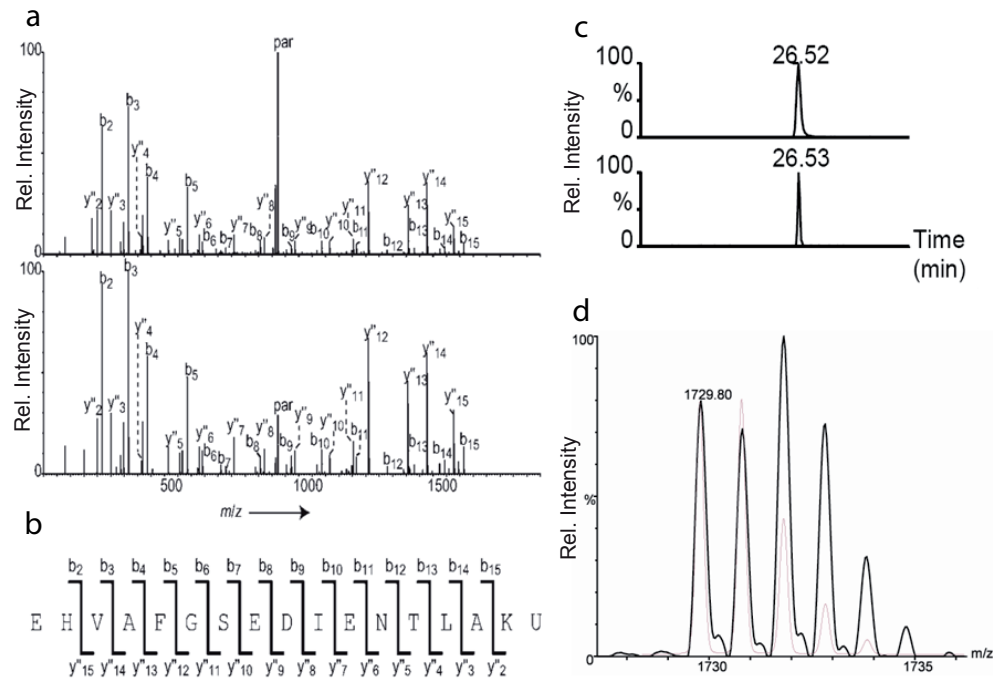

FIGURE 3. **Cleavage by TCEP results in a C-terminal homoserine lactone moiety.** Comparison of the cleavage of azPYP with TCEP and methionine containing PYP with cyanogen bromide, charge-deconvoluted tandem-MS fragmentation spectra of the doubly charged ion of m/z 921.9 that eluted at identical retention-times 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.

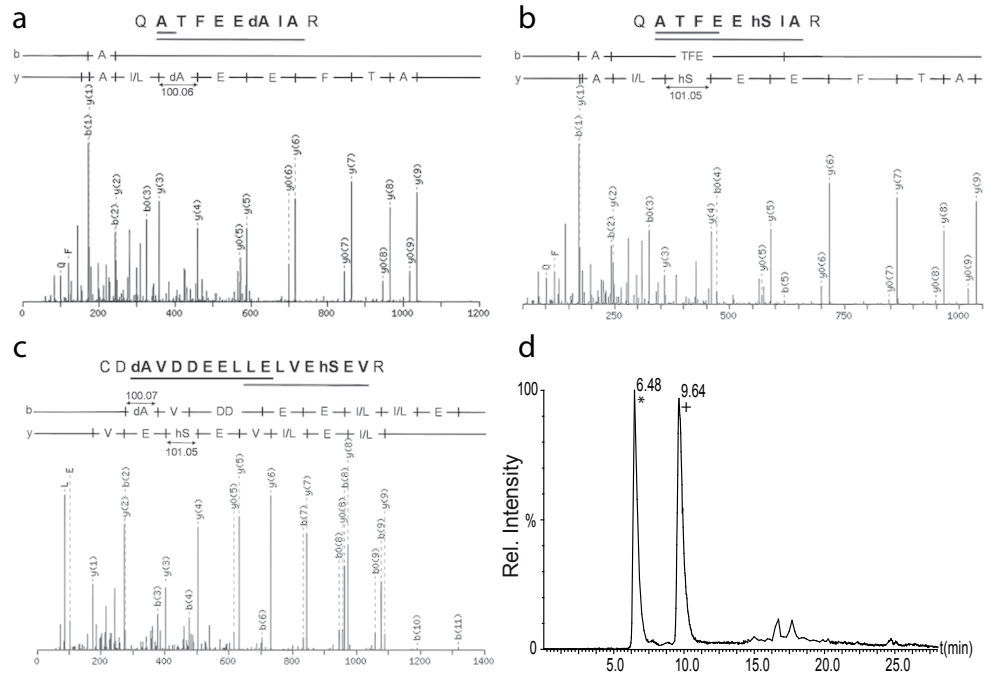

FIGURE 4. **The homoserine product is localized at the azhal-residue and has a different shift in retention-time compared to di-aminobutyrate.** Tandem-MS spectra of a peptide observed in two forms**.** The azhal residue can be reduced, forming di-aminobutyrate (dA), or react to form homoserine (hS) 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

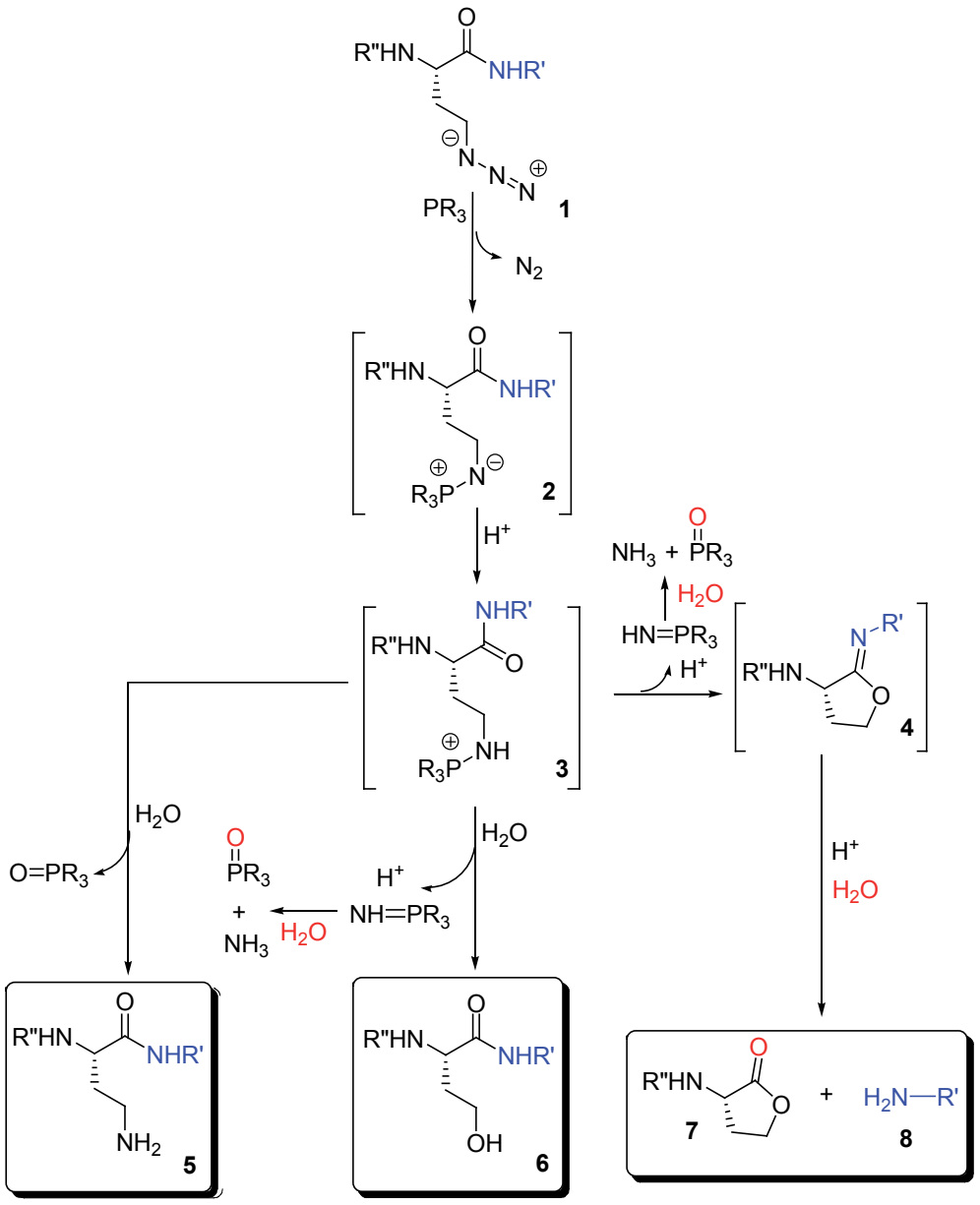

FIGURE 5. **Proposed reaction scheme of reduction and cleavage of azhal containing peptides by TCEP.** Scheme modified from Back *et al.* (143), PR_3=TCEP, which induces competing reactions on azhal-labelled peptides (1). The homoserine product (6) is formed by nucleophilic displacement of the iminophosphorane (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

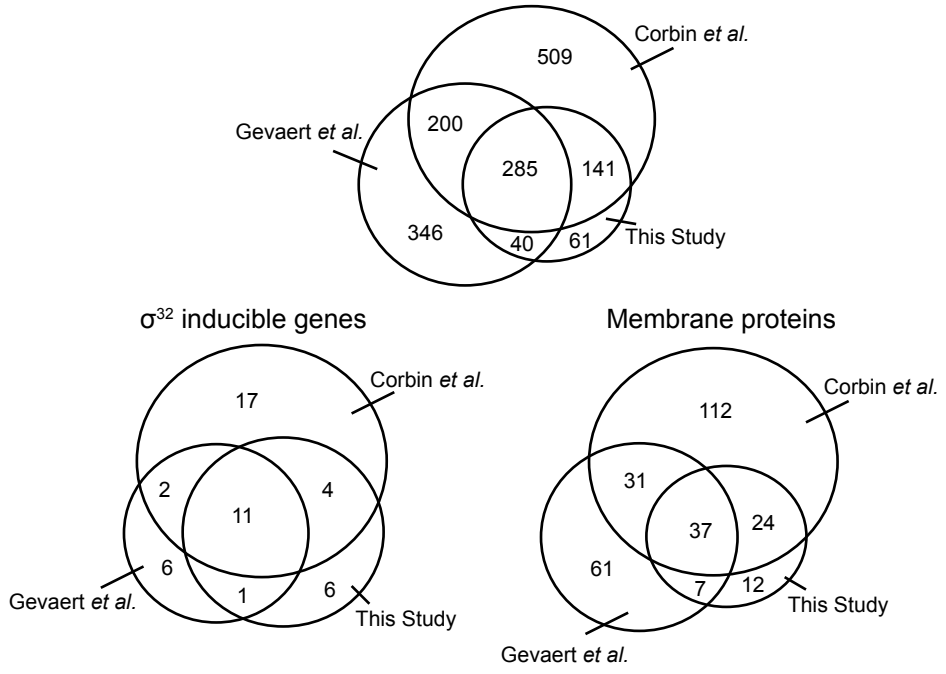

FIGURE 6. **Venn diagrams of proteins identified in this and alternative studies.** Overlap between proteins identified in this study and earlier studies of the *E. coli* proteome using 2D-LC by Corbin *et al.* (224) and COFRADIC, Gevaert *et al.* (213). Overlap between membrane proteins as well as σ^{32}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