2
LC–MS-based Protein and Peptide Quantification Using Stable Isotope Labels: From ICAT in General to Differential N-terminal Coding (dNIC) in Particular

ANDREAS O. WEINZIERL AND STEFAN STEVANOVIĆ*

Department of Immunology, Institute for Cell Biology, University of Tübingen, Auf der Morgenstelle 15, 72076 Tübingen, Germany

Introduction

Detailed quantitative comparisons of different cellular states are a key prerequisite for obtaining a firm understanding of the molecular basis behind various complex cellular pathways. The molecular mechanisms of interest can range from those underpinning regulatory consequences of a triggered signal cascade to those responsible for malignant cellular behaviour, such as in tumours. Different levels of quantitative comparisons of cellular states relevant to these and other processes have now become established over the past few years.

Regarding the so-called 'transcriptome', quantification has been well established for many years (Schena *et al.*, 1995). Single gene products can be quantified using qRT-PCR and even the complete entity of different transcriptomes are now quantitatively comparable in one single gene chip experiment. But, due to poor correlation

*To whom correspondence may be addressed (stefan.stevanovic@uni-tuebingen.de)

Abbreviations: 2D-GE, two-dimensional gel electrophoresis; ALICE, acid-labile isotope-coded extractants; AQUA, absolute quantification; cICAT, cleavable isotope-coded affinity tag; DIGE, difference gel electrophoresis; dNIC, differential N-terminal isotope coding; EDT, ethanedithiol; EST, expressed sequence tag; FARSLA, homologue to the phenylalanine-tRNA synthetase alpha unit; HLA-B*2704, human leukocyte antigen B*2704; ICAT, isotope-coded affinity tag; ICPL, isotope-coded protein labels; ICROC, isotope-coded reduction off a chromatographic support; LC, liquid chromatography; LCL721.174, lymphoblastoid B cell line 721.174; LCL721.45, lymphoblastoid B cell line 721.45; LC–MS, liquid chromatography–mass spectrometry; MALDI–TOF, matrix-assisted laser desorption ionization–time of flight; MHC-I, major histocompatibility complex class I; MS, mass spectrometry; MS/MS, tandem mass spectrometry; NIC, nicotinic acid; NIC-NHS, nicotinoyloxy-succinimide; OD, optical density; PhIAT, phosphoprotein isotope-coded affinity tags; qRT-PCR, quantitative real time PCR; SILAC, stabile isotope labelling by amino acids in cell culture; TAP, transporter associated with antigen processing; TMAB, 4-trimethylammoniumbutyryl.

Biotechnology and Genetic Engineering Reviews – Vol. 23, December 2006
0264–8725/07/23/21–39 $20.00 + $0.00 © Lavoisier/Intercept, 14 rue de Provigny, F-94236 Cachan cedex, France

between the transcriptome and the proteome (Gygi *et al.*, 1999a; Huber *et al.*, 2004), no general statements can be deduced from quantitative mRNA results in respect to changes in protein expression. In addition to the lack of correlation between the amount of an mRNA species and its corresponding protein, mRNA expression analyses are blind to post-transcriptional modifications of proteins (Ideker *et al.*, 2001). Thus, until now, a molecular analysis of a particular cellular state has had to be carried out at the proteome level. The higher chemical complexity of mixed protein systems makes the proteome not as easily quantifiable as the transcriptome but, within the past decade, several approaches have made quantitative comparisons between proteomes possible by high-throughput experimentation (Aebersold and Mann, 2003).

In general, all these approaches dealing with protein quantification have to solve several general problems. Firstly, the complexity of the proteome has to be reduced to a size that can feasibly be analysed, at best without losing valuable information and sensitivity. Secondly, the information gained in an experiment has to be interpreted by high-throughput analysis. Finally, the method has to be reproducible so as to be able also to detect minor changes in protein levels.

2D gel electrophoresis-based quantification techniques

The first attempt at protein quantification that met many of the above prerequisites was two-dimensional gel electrophoresis (2D-GE) (Klose, 1975; O'Farrell, 1975). Here, the complexity of a protein mixture is reduced by isoelectric focusing and subsequent size separation of the proteins. After separation of the proteins, they are stained in a gel, then isolated, enzymatically digested, and finally identified by MALDI–TOF-based peptide mapping (James *et al.*, 1993; Yates *et al.*, 1993). Although 2D-GE coupled to MS analysis is, in general, a valuable approach in quantitative proteome analysis, the classical quantitative 2D-GE has some major drawbacks. The lack of reproducibility between different gels hampers the distinction between variations of the sample and the system. The sensitivity of the system is limited, e.g. due to the type of protein dye chosen for protein visualization and also to limits in protein recovery from the gel (Gygi *et al.*, 2000). A major improvement in quantitative 2D-GE was achieved by Unlu and co-workers (Unlu *et al.*, 1997). These researchers introduced '*difference gel electrophoresis*' (DIGE), facilitating the analysis and comparison of several samples in a single 2D gel with increased sensitivity and reproducibility. In this approach, up to three samples are pre-labelled with spectrally different fluorescence dyes; the samples are then mixed and separated in the same gel. Consecutive 'spot' detection is then carried out with a scanner after laser excitation of the fluorescence dye. However, protein identification is still carried out after isolation and enzymatic digestion of each spot by MALDI–TOF. A major issue with 2D-GE proteomics that remains, therefore, is that of high-throughput protein spot selection and analysis. At least for gels stained with Coomassie Blue this can, in principle, be automated using robots (Yin *et al.*, 2004), although both accuracy and sensitivity are still key issues.

LC–MS-based quantification techniques

In order to circumvent difficulties arising from the 2D gel electrophoretic separation

of proteins, alternative quantitative proteomic techniques have been developed. Common to all of these techniques is the replacement of gel electrophoresis by liquid chromatography (LC) for protein separation and reduction of complexity. Generally speaking, the proteins in all techniques are enzymatically digested prior to separation, peptides are separated chromatographically and directly analysed

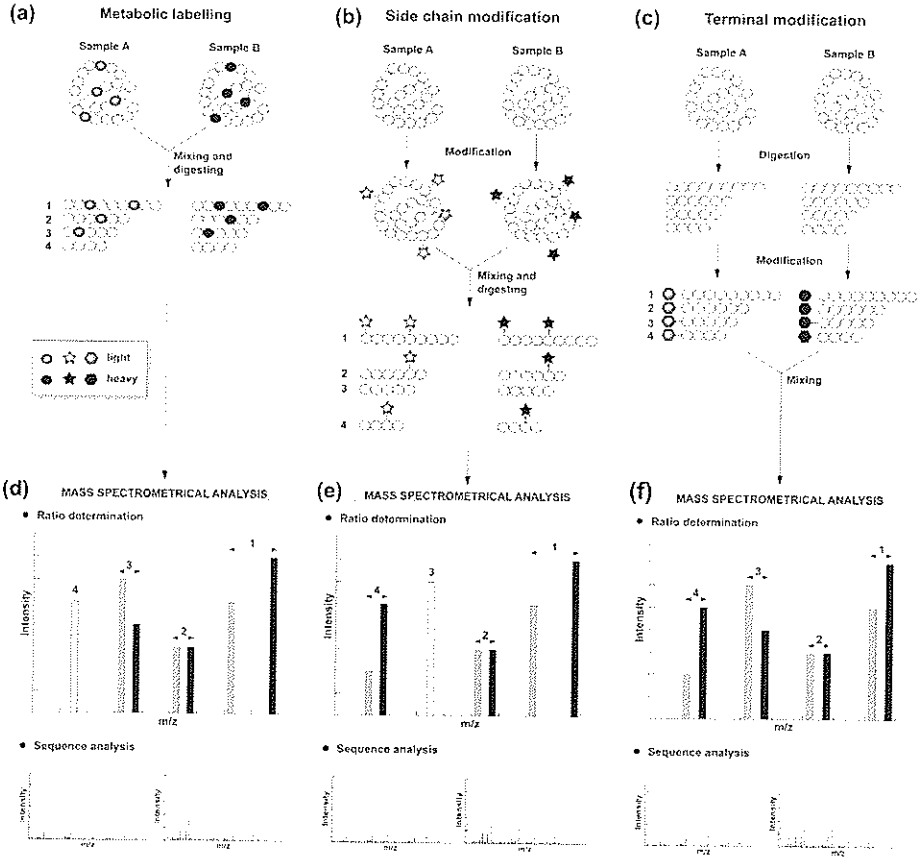


Figure 2.1. Schematic representation of stable-isotope labelling sites for protein labelling in quantitative proteomics. Light isotope marks are shaded in grey, heavy isotope marks in black, and unmarked amino acids are white. (a) and (d) Metabolic labelling: proteins are labelled metabolically by incorporation of ^{15}N, ^{13}C or isotope-labelled amino acids present in the cell culture medium. (b) and (e) Side chain modification: proteins are labelled at specific side chains with isotope-containing reagents. The reagents can also contain affinity tags for selective enrichment of tagged proteins. (c) and (f) Terminal modification: proteins are labelled C- or N-terminally using, e.g. enzymatic cleavage in $H_2{}^{18}O$ or chemical modification at the N-terminus, with isotope-containing reagents. Metabolic labelled samples and samples with side chain modifications are digested after isotope labelling, whereas for terminal modification, the samples are labelled during or after digestion. Afterwards, samples are quantitatively analysed by mass spectrometry (d–f, ratio determination). For metabolic labelling and side chain modification, various differences in the distance between light and heavy modified peptides (d and e), as well as unmodified peptides (d, peptide 4 and e, peptide 3), can be observed. Terminal modification shows a more uniform peptide pattern as no unmodified peptides emerge and distances between light and heavy modified peptides are equal (f). In a final step, peptides are sequenced by tandem MS/MS experiments (d–f, sequence analysis).

with an on-line coupled mass spectrometer (LC–MS). At some point during this pathway – depending on the approach selected – the peptide mixture ideally needs to be differentially labelled with isotope-coded tags. This approach is advantageous in several respects. On the one hand, peptide separation and data acquisition are automated *per se*, and therefore high-throughput analysis is easily possible. On the other hand, the method is highly reproducible, and therefore sensitive, even to minor changes in protein levels.

The fundamental obstacle in quantification by mass spectrometry is the fact that the signal yielded by a peptide in an MS experiment does not correlate to its total amount, and furthermore, the measurement is not accurately reproducible in consecutive experiments. This is true for MS experiments, irrespective of the ionization technique or the MS device used (Lill, 2003). The ionization efficiency of a specific peptide depends on the peptide sequence and on the nature and amount of co-eluting peptide features, which are not necessarily constant, especially when analysing quantitative differences between proteomes. This obstacle is circumvented by using isotope-coded labels that differ in mass but not in their physico-chemical behaviour on an LC column (Zhang *et al.*, 2002). Therefore, both species of a differentially-labelled peptide elute at the same retention time from the chromatography column, but are distinguishable in the mass spectrum recorded due to their different isotope-coded labels. In this case, the ionization efficiency for both peptide species is equal, and relative quantification is possible. This approach can also be used principally for absolute quantification (AQUA) (Gerber *et al.*, 2003), provided that one isotopically-labelled peptide species is a spiked peptide of known amount.

Practically, the basic principle of the procedure as outlined above comes in several variants, differing mainly in the kind of label itself and in the place of label introduction (*Figure 2.1a–c*). In the following section, the various isotope-coded modification reagents that have been developed in recent years are discussed. Then, we consider in detail examples of quantitative results using the particular technique of differential N-terminal isotope coding (dNIC).

METABOLIC LABELLING

Metabolic labelling of proteins occurs during protein biosynthesis within the cell (*Figure 2.1a*). Therefore, cells have to be grown in media enriched in the stable isotopes ^{15}N (Oda *et al.*, 1999; Washburn *et al.*, 2002) or ^{13}C (Stocklin *et al.*, 2000), or in media which are supplemented with isotope-incorporated amino acids, such as D_3-leucine (Ong *et al.*, 2002) and arginine with various levels of ^{13}C and ^{15}N, e.g. $^{13}C_6$-Arg, $^{13}C_6$-$^{15}N_4$-Arg (Andersen *et al.*, 2005) or $^{15}N_2$-Arg (Ringrose *et al.*, 2004) (*Figure 2.2*).

The use of ^{15}N-enriched media has been found to provide an accurate quantification method and is more frequently used than ^{13}C-modified media. In order to reduce difficulties and expense in preparing ^{15}N-enriched media for mammalian cells, the use of stable isotope labelling by amino acids in cell culture (SILAC) for differential proteomics was introduced in 2002 by Ong and colleagues (Ong *et al.*, 2002). Several quantitative analyses have been carried out ranging from an analysis of proteome dynamics of the nucleolus (Andersen *et al.*, 2005) and proteome changes

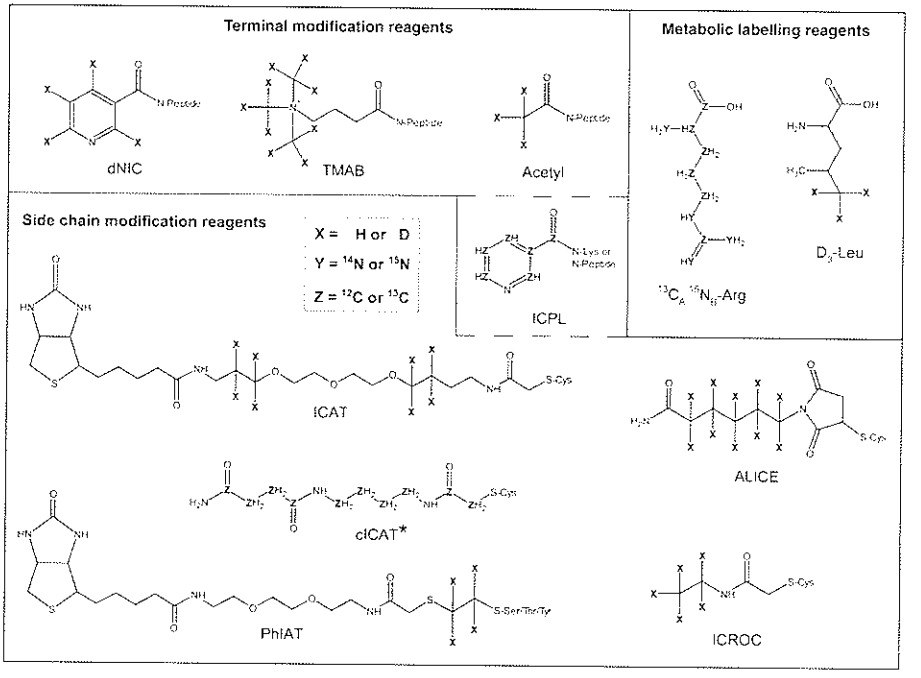


Figure 2.2. Structure formulae of isotope-bearing modification reagents described in this review. Sites bearing heavy isotopes in the respective heavy isotope reagents are marked with X, Y, and Z. For further details regarding the reagents, see the respective sections of the text. *The heavy cICAT reagent contains nine ^{13}C-atoms out of the ten C-atoms in total. The exact position of these nine ^{13}C-atoms is a corporate secret of Applied Biosystems, Framingham, MA, USA.

in muscle cell differentiation (Ong *et al.*, 2002) to repertoire changes of the major histocompatibility complex class I (MHC-I) peptides after *Salmonella enterica* infection (Ringrose *et al.*, 2004).

A principal obstacle concerning metabolic labelling is that the number of heavy atom-containing amino acids incorporated into one peptide is not deducible without sequence information. Mass differences between isotopic labelled peptides depend on the sequence varying from peptide to peptide (*Figure 2.1a,d*: compare peptides 1–4). An additional restraint is the fact that highly abundant peptides leading to just one MS signal are indistinguishable from any unlabelled peptides (*Figure 2.1d*, peptide 4). This in turn dramatically reduces the impact of quantifiable MS data if sequence information is not available.

CHEMICAL LABELLING

For proteomic samples that cannot be derived from *in vitro* cultured cells, the metabolic labelling approach is not applicable. These samples can be chemically modified with isotope labels at their terminus or at side chains. This chemical modification can be divided into side chain modifications (*Figure 2.1b,e*) and terminal modifications (*Figure 2.1c,f*). The major difference between side chain modification and terminal modification is not only the modification reagent (see

Figure 2.2 dNIC and ICPL) but also the number of peptides available for modification, and thus for quantification. After side chain labelling, only peptides containing the targeted amino acid are modified, and the number of labels per peptide is also sequence dependent (*Figure 2.1b,e*, compare peptides 1–4), whereas terminal labelling introduces exactly one label in every peptide (*Figure 2.1c,f*: compare peptides 1–4).

SIDE CHAIN LABELLING

Favourite targets for specific modifications of side chains are the amino group of lysine residues or the thiol group of cysteines. In principle, proteins are first isolated, then modified and digested enzymatically, and subsequently subjected to MS analysis (*Figure 2.1b*).

ICAT and cICAT

In 1999, Gygi and colleagues introduced the first isotope-coded affinity tag (ICAT) (Gygi *et al.*, 1999b). The ICAT reagent consists of a biotin tag, an isotope carrying linker, and a reactive sulfhydryl group for the attachment of the ICAT reagent to reduced cysteine residues. The biotin tag allows the purification of labelled peptides, reducing the sample complexity on the one hand and omitting singular signals for the analysis of differences in the *Saccharomyces cerevisiae* proteome induced by galactose or ethanol as carbon source on the other (Gygi *et al.*, 1999b); the ICAT reagent itself has some restraints. The biotin isotope tag is relatively large (442 Da) and can result in incomplete peptide sequence information (Lill, 2003). Furthermore, the purification via the biotin–avidin interaction is susceptible to non-specific binding and incomplete elution (Moseley, 2001), and finally, the eight deuterium atoms of the heavy ICAT reagent were seen to cause a relatively strong isotope effect on the chromatography column, which could partially resolve isotopic peptide pairs (Tao and Aebersold, 2003). These problems were largely resolved with several improvements to the ICAT reagent. The cleavable ICAT (cICAT, *Figure 2.2*) has two major improved features. Firstly, it contains an acid-cleavable linker, which allows removal of the biotin tag from the labelled peptide. This removal of the biotin decreases the mass of the ICAT for MS analysis and results in better interpretable peptide fragmentation spectra. Additionally, the isotope effect in chromatography has now been omitted by replacement of the deuterium with ^{13}C-isotope labels (Hansen *et al.*, 2003).

ICROC and ALICE

Several other strategies vary the basic idea of ICAT modification. For isotope-coded reduction off a chromatographic support (ICROC, *Figure 2.2*), proteins are also reduced and cleaved with trypsin (Shen *et al.*, 2003). In this way, the cysteine-containing peptides thus generated are covalently coupled to pyridyl disulfide beads. After washing, the cysteine peptides are eluted using a reducing agent. Following this, the free cysteine thiol groups are alkylated with either N-ethyl or N-D_5-ethyl-iodacetamide and subjected to LC–MS for quantitative analysis.

The use of acid-labile isotope-coded extractants (ALICE, *Figure 2.2*) (Qiu *et al.*, 2002) also circumvents the need for biotin/avidin-based purification of cysteine peptides. Cysteine-containing peptides derived from tryptic digests are linked via their thiol groups to the ALICE reagent, which, in addition to a thiol group, also contains an isotope-coded linker and an acid labile unit, covalently coupled to an inorganic resin. After linking the cysteine peptides to the resin, they are washed, eluted by acid treatment, and analysed sequentially.

PhIAT

The principle of ICAT has been conveyed also to the quantitative analysis of phosphorylated proteins using phosphoprotein isotope-coded affinity tags (PhIAT). In 2001, Oda and colleagues developed a labelling strategy for phosphorylated proteins with a reagent resembling ICAT (Oda *et al.*, 2001). Proteins were subjected to tryptic digestion, thiol groups of cysteine residues were oxidized to cysteic acid and thus blocked from further reactions. Phosphate residues were removed by β-elimination using high pH conditions. This elimination results in a replacement of the phospho group with an unsaturated residue to which ethanedithiol (EDT) is added. Finally, a biotin-bearing linker structurally related to the ICAT linker is introduced to the free thiol group of the EDT. The additional steps of processing are the same as those for the ICAT labelled peptides considered above, namely biotin-based purification, followed by subsequent elution and LC–MS analysis. This approach is easily applicable for quantification using EDT-H_4 and EDT-D_4 for the induction of isotope-bearing tags (*Figure 2.2*) for peptide modification (Goshe *et al.*, 2002).

ICPL

The purification of cysteine peptides to reduce the complexity of the peptide mixture by some 10-fold is a major advantage of the ICAT method, but can also become one of its pitfalls. The reduction of complexity leads also to a reduction of protein coverage, as cysteine is one of the rarest amino acids. For example, approximately 66% of all *E. coli* open reading frames contain less than five cysteine residues (Schmidt *et al.*, 2005).

Apart from thiol groups occurring in cysteine residues, the more frequently occurring amino groups in, for example, lysine residues are also open to modification in a selective way. More than 80% of all proteins have a high lysine content and are therefore available for isotope-coded protein labelling (ICPL, *Figure 2.2*) using either D_4-nicotinic acid (Schmidt *et al.*, 2005) or $^{13}C_6$-nicotinic acid (Hochleitner *et al.*, 2005) as the modifying reagent. Similar to cysteine-based quantification approaches, in the case of lysine, the proteins are first modified, then mixed and digested. As lysine modification impairs protein digestion with trypsin, proteins have to be digested with other proteases, such as endoproteinase Glu-C.

Although the higher protein coverage of the ICPL technique can be very valuable, some limitations are intrinsic to this approach. Unlabelled peptide portions are not separated from the modified peptide pool, and thus singular peptide peaks in LC–MS analysis derive from highly abundant peptides as well as from unmodified ones.

Additionally, multiple labelling of one peptide is likely to be due to the fact that lysine is a frequent amino acid. This multiple labelling in turn complicates identification of peptide pairs (*Figure 2.1e*).

TERMINAL LABELLING

In principle, terminal labelling has one major advantage as every peptide is labelled, and thus all peptides are available for quantification. Protein labelling after digestion bears the restraint that mixing of samples occurs at a later stage, compared with all other approaches (*Figure 2.1c*), thus rendering it *per se* more prone to artifacts due to a separated sample preparation.

Enzymatic digestion in the presence of $H_2{}^{18}O$

One of the first strategies used for quantitative comparisons of protein samples was the enzymatic digestion of protein preparations in ^{18}O-labelled water, marking the C-terminus of the generated peptide by a mass shift of 2 Da compared to peptides digested in $H_2{}^{16}O$ (Stewart *et al.*, 2001; Yao *et al.*, 2001). Although this approach is elegant and fast, there are some restraints. The small change in the peptide mass due to the incorporation of ^{18}O results in an overlay of the isotope pattern of light and heavy tagged peptides. Furthermore, the 2 Da mass difference between differentially-tagged peptides can be altered due to rebinding of trypsin to peptidic arginine and lysine residues after cleavage. The rebinding leads to the incorporation of a second ^{18}O into the peptide (Reynolds *et al.*, 2002), complicating the quantitative analysis of the sample.

Acetylation and TMAB

Acetylation of free amino groups in a peptide is both chemically simple and

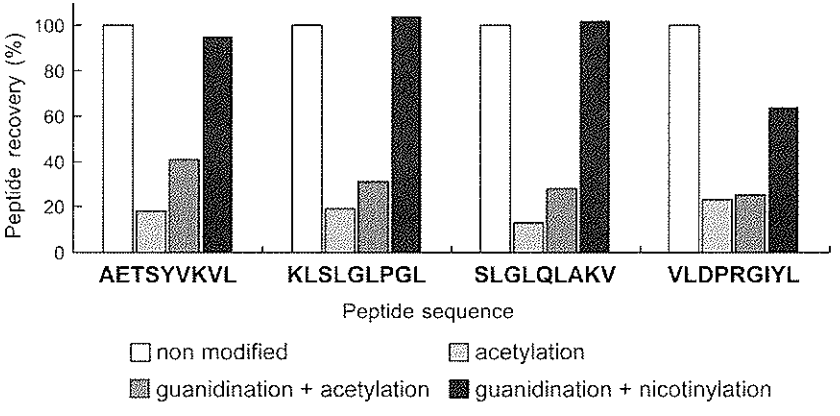


Figure 2.3. Recoveries of different peptide derivatives. The four synthetic peptides (AETSYVKVL, KLSLGLPGL, SLGLQLAKV, and VLDPRGIYL) present in equimolar concentrations were either acetylated, guanidinated and acetylated, or guanidinated and nicotinylated. After complete derivatization, each peptide mixture was combined with the initial non-modified peptide mixture (thus expecting equimolar yields of both species) and quantified by nanospray–ESI–MS analysis. This figure has been adapted, with permission, from Lemmel *et al.* (2004).

quantitative, hence easily facilitating the introduction of an isotope-coding tag into a peptide. Several protocols have been described for either the specific acetylation of the N-terminus after blocking free amino groups in lysine side chains (Lemmel *et al.*, 2004; see also dNIC section) or the non-specific acetylation of all free amino groups in a peptide (Che and Fricker, 2002), which is less favourable for analysis.

Irrespective of the specificity of the reaction, acetylation dramatically reduces the nucleophilic character of the modified amino group, resulting in highly ineffective ionization efficiencies of the acetylated peptides (*Figure 2.3*). Therefore, other N-terminal modification reagents, which preserve a positive N-terminal charge, are more desirable.

One example of such a reagent is 4-trimethyl-ammoniumbutyryl (TMAB, *Figure 2.2*), which also reacts with every free amino group (Zhang *et al.*, 2002). In 2005, Che and co-workers used this reagent without blocking lysine side chains for the successful quantification of neuropeptides in mice (Che *et al.*, 2005).

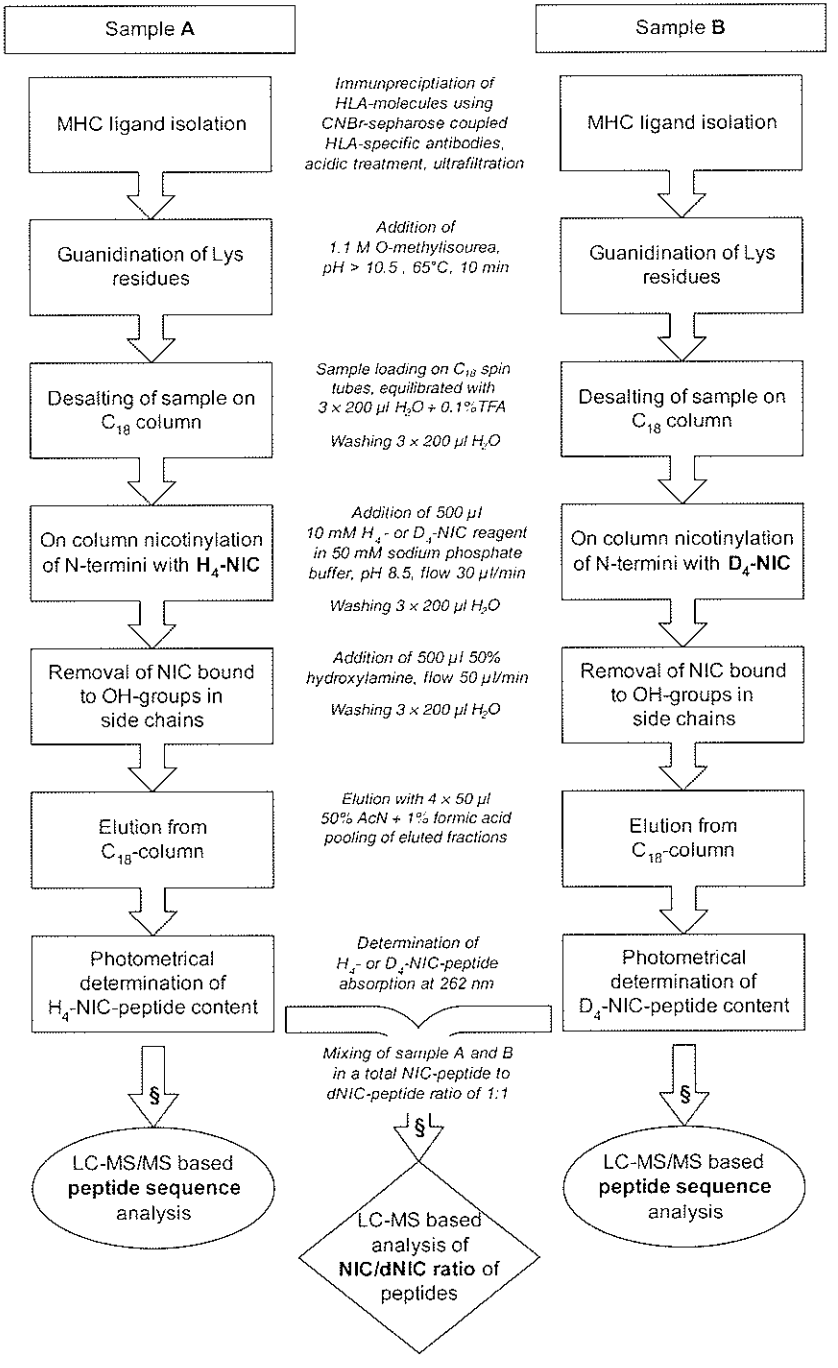
dNIC

In 2004, Lemmel and co-workers introduced an approach for providing a differential N-terminal isotope coding (dNIC) selectively at the N-terminus (Lemmel *et al.*, 2004). To prevent isotope labelling of lysine side chains, these amino groups are selectively guanidinated at a pH above 10.5 using O-methylisourea (Beardsley and Reilly, 2002): a workflow illustrating this process is shown in *Figure 2.4*. The peptides are desalted on C_{18}-columns and afterwards the dNIC label is introduced solely at the N-terminus. This approach is advantageous in several respects. Every peptide is modified with exactly one dNIC label, meaning that single LC–MS events cannot emerge from unlabelled peptides but can only arise from singularly existing peptides in one of the samples. The dNIC label is very light (110 Da) and thus does not have the same problems incurred by the use of ICAT labels in respect of peptide sequencing. Furthermore, the dNIC label, as well as the guanidination process, enhances MS sensitivity (Keough *et al.*, 2000; Brancia *et al.*, 2001), which quantitatively compensates for losses occurring during peptide labelling (*Figure 2.3*).

Peptide quantification using dNIC

QUANTITATIVE ANALYSIS OF MHC-I LIGANDS

MHC-I molecules are cell surface proteins presenting a non-covalently bound peptide (Saper *et al.*, 1991). These MHC ligands, which mainly derive from endogenous proteins, are displayed primarily to cytotoxic T cells for immunosurveillance (Zinkernagel and Doherty, 1974). Several features of MHC ligands exclude them from ICAT-based quantification but make them perfectly suited for N-terminal modification using the dNIC approach. The MHC-bound peptides range in size from 8–10 amino acids (Rammensee *et al.*, 1993) and seldom contain cysteine residues (as can be seen from the MHC ligand database at www.syfpeithi.de). Thus, all ICAT-based approaches are not suited for quantitative MHC analysis. Fortunately, the complexity of an MHC ligand preparation is



§ *Prior to LC-MS or LC-MS/MS analysis AcN was removed from samples by vaccum centrifugation. Samples were resuspended in H_2O containing 4 mM ammonium acetate, adjusted to pH 3 with formic acid*

Figure 2.4. Workflow for dNIC-based quantitative comparison of MHC-I ligand levels isolated from two samples.

sufficiently small so as to negate the need for reduction prior to LC–MS analysis. Ringrose and co-workers have therefore used metabolic labelling with ^{15}N-arginine to analyse differential HLA-B*2704 peptide presentation after *Salmonella enterica* infection (Ringrose *et al.*, 2004). The HLA-B*2704 molecule binds mostly peptides which present in their second position an arginine as a so-called 'anchor residue'. In this way, all isolated HLA-B*2704 ligands were differentially marked at position two. This approach is not generally conferrable to other MHC ligand pools as most MHC motifs are not restricted to one specific amino acid in their anchor positions in contrast to the arginine anchor in HLA-B*2704 presented peptides. A metabolic labelling with a more abundant amino acid would generate the same problems as described above for metabolic labelled proteins. Due to the fact that protein coverage of MHC-bound peptides is very low *per se*, it is, in this context, particularly important that virtually every peptide is quantitatively analysable. Therefore, a quantification based on lysine side chain labelling, such as ICPL, which has no separation of labelled peptides from unlabelled peptides, would also be inadequate.

All these restraints, which hamper metabolic and side chain labelling, do not apply to the dNIC approach (*Figure 2.4*), as it is independent of amino acid composition, with all peptides available for quantification (*Figure 2.1c*).

KINETICS OF MHC-I LIGAND NICOTINYLATION

Peptide modification, either at the amino groups of lysine side chains or at the N-terminus with N-nicotinoyloxy-succinimide (NIC-NHS) esters, is performed from between 1 h (Hochleitner *et al.*, 2005) and 2.5 h (Munchbach *et al.*, 2000; Schmidt *et al.*, 2005). In order to optimize reaction time and NIC-NHS usage, we have investigated nicotinylation kinetics in a pulse-chase experiment (*Figure 2.5*). A mixture of

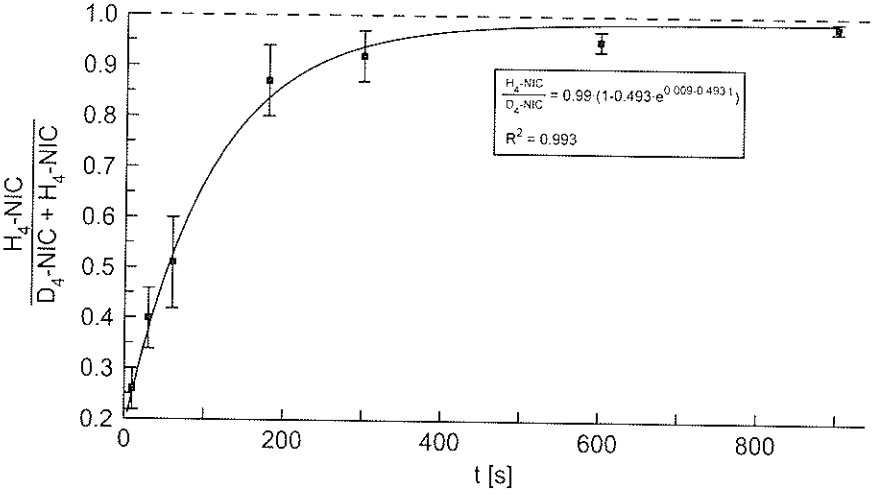


Figure 2.5. Kinetic analysis of peptide nicotinylation. A mix of eight synthetic peptides was guanidinated and afterwards subjected to nicotinylation for 15 min in total. First, peptides were pulsed with 10 mM H_4-NIC for different times indicated in the diagram. Afterwards, H_4-NIC was removed and, for the remaining time, 10 mM D_4-NIC was added for chase. For each peptide and point in time, ratios of peptides modified with light and heavy nicotinic acid were calculated. The fitting function employed is indicated in the diagram.

eight synthetic peptides (4 nmol each peptide) was (on-column) modified within 15 min in total, with 10 mM of light or heavy nicotinylation reagent at a flow rate of 35 µl/min. For pulsing, the pre-guanidinated peptide mix was incubated for 10 s to 15 min with H_4-NIC. The chase was carried out by replacing light nicotinylation solution with D_4-NIC to give a total nicotinylation time of 15 min. For kinetic analysis, the H_4-NIC to D_4-NIC ratio of modified peptides was determined by mass spectrometry. Summarizing the results, a nicotinylation efficiency of 99% was achieved for every synthetic peptide after 15 min, and thus in this way, the protocol for the dNIC approach could be optimized.

APPROPRIATE MIXING OF dNIC-LABELLED MHC LIGANDS

Adequate mixing of the samples prior to quantification analysis is one of the most crucial steps for every quantitative application. Regarding adequate mixing, metabolic labelling is generally advantageous, as with no other isotope labelling method can different samples be mixed so early in the workflow (*Figure 2.1a*). For quantitative analysis, total cell preparations can be normalized to equal cell number or total protein content, directly mixed, and processed together. This early mixing procedure reduces any artificial protein quantification differences resulting from variations in sample handling. Principally, terminal modification implies a later mixing than all other modification techniques, but this does not apply to MHC-I ligand samples. MHC-I ligands are short peptides, therefore an enzymatic digestion step is surplus. Nevertheless, accurate determination of total MHC ligand content is difficult. Before acid treatment of immunoprecipitated MHC, the total amount of MHC can be determined, and hence the quantity of the isolated MHC ligands thereof can only be roughly deduced.

Due to the sensitivity of LC–MS-based dNIC quantification experiments, a more accurate determination of total peptide content of a sample prior to mixing is required. As peptide nicotinylation introduces a strongly UV-absorbing nicotinic acid into every peptide, the total amount of peptide in a sample can be determined by UV-absorption at 262 nm. To prove that this densitometric-based mixing is, in principle, appropriate for LC–MS quantification experiments, pools of synthetic peptides have been differentially labelled and mixed in fixed ratios according to their OD_{262nm}. Afterwards, the actual peptide ratios were determined by MS. In detail, 6, 12, and 24 nmol of six synthetic peptides were modified, either with light or heavy nicotinic acid. After elution from the C_{18} columns, the OD_{262nm} values were determined for each sample, whereas an equally treated sample without peptide served as the blank. According to their optical densities, the samples were mixed in two ratios of 1:1 and 0.8:1, respectively. The actual peptide ratios were determined by mass spectrometric analysis (*Table 2.1*). The results from this experiment demonstrate that UV absorption-based peptide mixing is accurate for nicotinylated peptides, as the relative error was as low as 8%, which is in a range similar to that obtained with other isotope-based quantification strategies (Gygi *et al.*, 1999b).

This mixing method was also successfully applied to more complex samples, such as total MHC ligands from tumour and normal tissue, modified with D_4-NIC or H_4-NIC, respectively (*Figure 2.6*). Here, most MHC ligands seem to be equally presented on tumour and autologous normal tissue. Assuming a Gaussian distribution of the

Table 2.1 Proof of principle for densitometry-based mixing of peptides modified with the dNIC strategy. 42 nmol of a peptide mix consisting of the peptides indicated were guanidinated together and split into portions of 6, 12, and 24 nmol. These portions were separately modified with H_4-NIC (6 and 24 nmol portion) and D_4-NIC (12 nmol portion). The optical density at 262 nm was determined for each sample after modification. (a) The 6 nmol and 12 nmol portions were mixed according to their optical density in a 1:1 ratio. (b) The 12 and 24 nmol portions were mixed in a 0.8:1 ratio, respectively. The actual peptide ratios were determined by MS; average and standard deviation of the measured ratios were calculated.

Amount of modified peptide	6 nmol	12 nmol		24 nmol	12 nmol	
Peptide sequence	**Intensity**		**ratio**	**Intensity**		**ratio**
	NIC	**dNIC**		**NIC**	**dNIC**	
PGSYTYEWNFRKDVN	386	393	0.98	612	755	0.81
DVETQFNQYK	547	506	1.08	966	1217	0.79
KLKEFIPKV	495	460	1.08	514	629	0.82
GLDEVKSSL	696	607	1.15	1021	1251	0.82
VVQLTLAFR	1683	1633	1.03	1777	2462	0.72
ILNSWNISK	799	703	1.14	1286	1444	0.89
Average			**1.08±0.06**			**0.81±0.05**
Expected			**1.00**			**0.80**

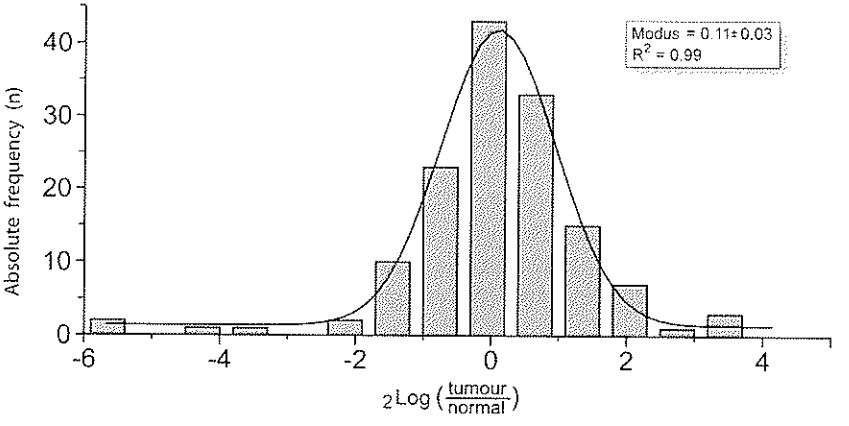


Figure 2.6. Densitometry-based mixing of two complex peptide samples after dNIC modification. MHC-I ligands were isolated from primary renal cell carcinoma tumour and autologous kidney normal tissue, modified with D_4-NIC or H_4-NIC, respectively, and mixed according to their OD_{262nm} in a 1:1 ratio. A frequency count (bin size 0.7) of the logarithmized ratios of tumour- to normal tissue-derived ligands was performed and fitted using a Gaussian curve; mode and R^2 are indicated within the figure.

logarithmized ratio of tumour- to normal tissue-derived peptides (R^2 = 0.99), the ratios can be further normalized setting the mode of the curve at zero.

APPLICATION OF dNIC: QUANTIFICATION OF TAP INDEPENDENTLY PRESENTED PEPTIDES

The dNIC-modification strategy (*Figure 2.4*) was used for analysis of quantitative comparisons of cell lines with defects in their antigen processing due to transporter associated with antigen (TAP) deletion. For this experiment, equal amounts of MHC ligand complexes were isolated from the $TAP^{-/-}$ cell line, LCL721.174, and from its

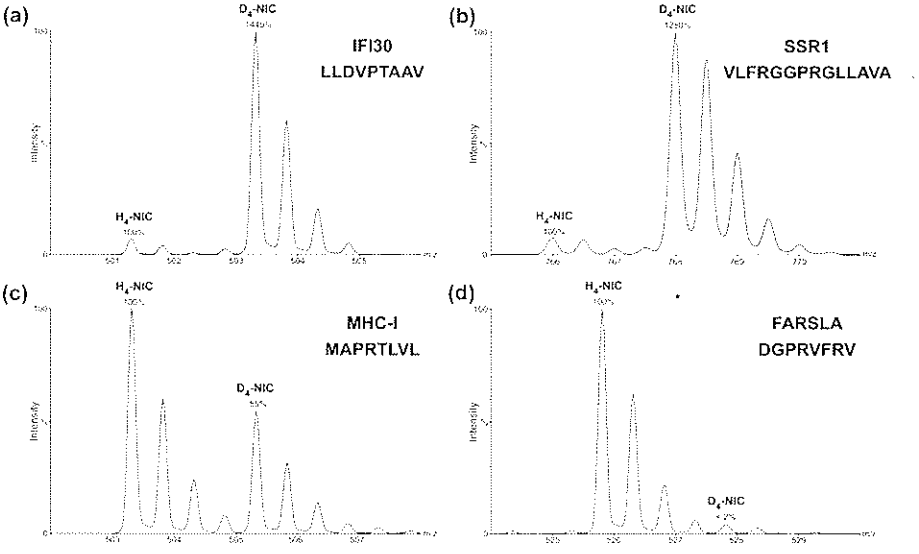


Figure 2.7. Ratio determination of different MHC-I ligands derived from TAP-expressing or TAP-deficient cells. From LCL721.45 ($TAP^{+/+}$) and LCL721.174 ($TAP^{-/-}$) cells, MHC-I ligands were isolated and modified with H_4-NIC or D_4-NIC, respectively. Modified peptide pools were mixed in a total peptide ratio of 1:1 and subjected to LC–MS analysis. For the peptides indicated, the relative amount of peptide derived from LCL721.174 was calculated.

TAP-expressing progenitor cell line, LCL721.45 (Koller *et al.*, 1989). Afterwards, peptides were eluted from the isolated MHC molecules. LCL721.174-derived peptides were modified with the heavy nicotinylation reagent, and LCL721.45 with the light reagent. MHC ligand ratios were mixed in a total peptide ratio of 1:1 by densitometry. Two peptides from IFI30 and SSR1 have been identified previously on LCL721.174 by Wei and Cresswell (1992). The presentation level of these two peptides was compared between the two LCLs (*Figure 2.7*). The IFI30-derived peptide, LLDVPTAAV (*Figure 2.7a*), and the SSR1-derived peptide, VLFRGGPRGLLAVA (*Figure 7b*), were both more than 12-fold over-presented on MHC-I molecules of LCL721.174. Due to quite different MHC ligand pools on both LCLs, normalization using a Gaussian fitted ratio of LCL721.174- to LCL721.45-presented peptides (compare with *Figure 2.6*) was not possible. It is notable that peptides derived from MHC signal sequences are abundantly presented on MHC molecules. The signal-derived peptide, MAPRTLVL, was found to be approximately twofold under-presented on the LCL721.174 (*Figure 2.7c*). This suggests that the IFI30- and SSR1-derived peptides were not over-represented due to inadequate mixing of the samples but reflect a biological over-presentation.

As an example of a peptide whose MHC-I presentation is TAP dependent, *Figure 2.7d* shows the ratio of the peptide, DGPRVRFV, derived from the alpha unit of a homologue to the phenylalanine-tRNA synthetase (FARSLA). The isotopic pattern of this peptide derived from LCL721.45 overlays with the signal of the potential peptide from LCL721.174. But the identical experiment using switched modifications between the two peptide pools suggests that this peptide is not presented on LCL721.174 (<0.5%, data not shown).

In this exemplary dNIC quantification experiment, the observed dynamic range of dNIC-based quantification was about three orders of magnitude, a range that concurs with other isotope-based LC–MS quantification experiments (Gygi *et al.*, 1999b).

AUTOMATIZATION OF PEPTIDE QUANTIFICATION USING dNIC LABELS

The dNIC strategy labels every peptide and introduces a mass shift from light to heavy modification reagent of 4 Da, independent of the peptide sequence. This constant mass difference, and the isotopic effect of the peptides modified with the deuterium-bearing nicotinic acid, can be used for reliable automated analysis of peptide pair ratios. This makes the dNIC quantification approach well suited for high-throughput experiments.

FACILITATED *DE NOVO* SEQUENCING BY dNIC LABELLING

Apart from the fact that N-terminal modification with heavy and light nicotinic acid enables quantitative comparison of two samples, it also facilitates *de novo* peptide sequencing. Identical peptides modified with light and heavy dNIC reagent result in identical MS/MS spectra, except that every fragment containing the N-terminus differs in mass by 4 Da. Taking advantage of that fact, we could *de novo* sequence the MHC ligand, IITKEVLAP, which origins only in an expressed sequence tag (EST) corresponding to a yet unknown protein (*Figure 2.8*). The modified synthetic peptide (*Figure 2.8a*) displays an identical fragmentation pattern compared to the MHC ligands isolated from the LCL721.174 cell line modified with light or heavy modification reagent (*Figure 2.8b/c*). Another feature of nicotinylation facilitating sequencing is the fact that the N-terminal b-ion (*Figure 2.8d*) becomes observable in MS/MS fragmentation spectra. Normally, N-terminal b1-ions (*Figure 2.8d*) are not generated during peptide fragmentation, as b-ion generation requires cyclic transition states (Yalcin *et al.*, 1995). Instead, b2-ions are generally well observable. As nicotinylation introduces a new peptide bond at the N-terminus, the b2-ions generated from nicotinylated peptides contain only nicotinic acid and the N-terminal amino acid, and thus make the N-terminal amino acid easily identifiable.

Conclusions

Although there is still much room for further improvements regarding, for example, quantification of very low abundant proteins or peptides and automation, quantitative comparisons of proteins using LC–MS has, during the past decade, clearly become a well-established technique.

Several different modification reagents cover a broad range of biological applications, from total proteome comparisons using ICAT or ICAT-linked techniques to reagents suited for phosphoprotein or MHC ligand quantification. Common to all these techniques is good reproducibility and sensitivity, as well as a broad dynamic range, all needed for accurate quantification experiments. Therefore, it can be expected that protein quantification will be used more frequently in the future, tackling multifaceted problems in various biological systems, and that it will

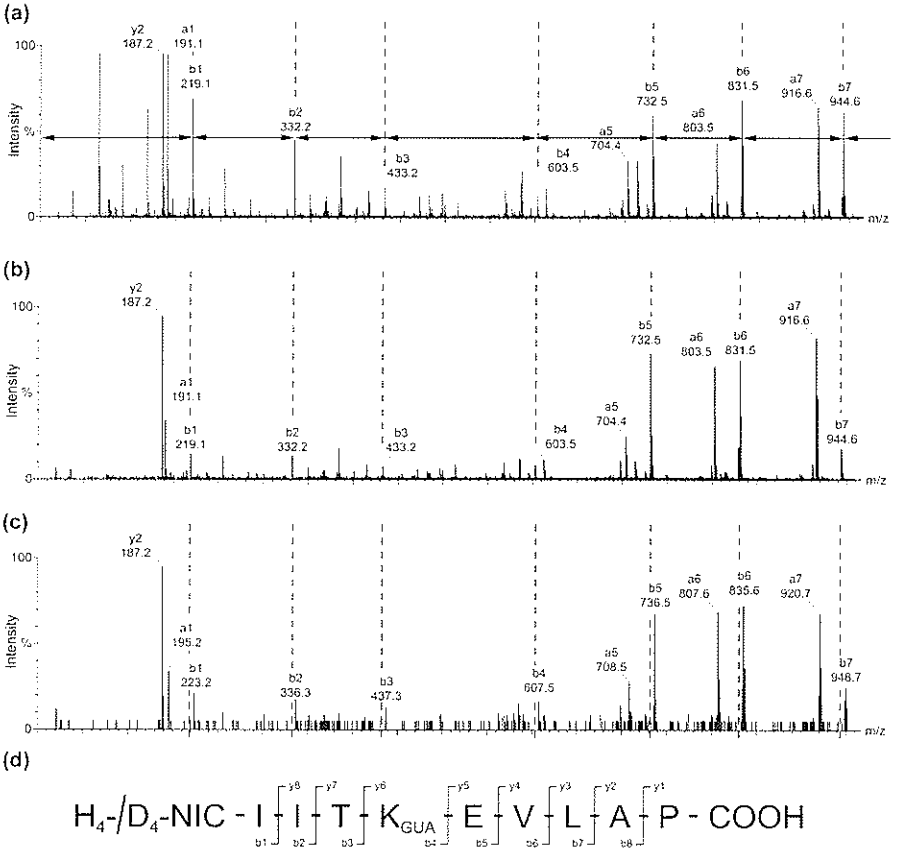


Figure 2.8. Facilitated *de novo* sequencing of the modified peptide IITKEVLAP using the dNIC approach. (a) MS/MS fragmentation spectrum of synthetic IITKEVLAP modified with D_4-NIC after guanidination. (b) MS/MS fragmentation spectrum of IITKEVLAP isolated from LCL721.174, also modified with D_4-NIC after guanidination. (c) MS/MS fragmentation spectrum of guanidinated and H_4-NIC modified IITKEVLAP, isolated from LCL721.174. (d) Nomenclature of MS/MS fragments derived from modified IITKEVLAP.

increasingly complement the large availability of quantitative mRNA data (Ideker *et al.*, 2001).

Acknowledgements

We thank Claudia Lemmel, who invented the dNIC approach, for invaluable help during optimization of the method, and Lynne Yakes for expert proof reading. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 685 and SFBTR 19).

References

AEBERSOLD, R. AND MANN, M. (2003). Mass spectrometry-based proteomics. *Nature* **422**, 198–207.

ANDERSEN, J.S., LAM, Y.W., LEUNG, A.K. *ET AL.* (2005). Nucleolar proteome dynamics. *Nature* **433**, 77–83.

BEARDSLEY, R.L. AND REILLY, J.P. (2002). Optimization of guanidination procedures for MALDI mass mapping. *Analytical Chemistry* **74**, 1884–1890.

BRANCIA, F.L., BUTT, A., BEYNON, R.J., HUBBARD, S.J., GASKELL, S.J. AND OLIVER, S.G. (2001). A combination of chemical derivatization and improved bioinformatic tools optimizes protein identification for proteomics. *Electrophoresis* **22**, 552–559.

CHE, F.Y. AND FRICKER, L.D. (2002). Quantitation of neuropeptides in Cpe(fat)/Cpe(fat) mice using differential isotopic tags and mass spectrometry. *Analytical Chemistry* **74**, 3190–3198.

CHE, F.Y., BISWAS, R. AND FRICKER, L.D. (2005). Relative quantitation of peptides in wild-type and Cpe(fat/fat) mouse pituitary using stable isotopic tags and mass spectrometry. *Journal of Mass Spectrometry* **40**, 227–237.

GERBER, S.A., RUSH, J., STEMMAN, O., KIRSCHNER, M.W. AND GYGI, S.P. (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 6940–6945.

GOSHE, M.B., VEENSTRA, T.D., PANISKO, E.A., CONRADS, T.P., ANGELL, N.H. AND SMITH, R.D. (2002). Phosphoprotein isotope-coded affinity tags: application to the enrichment and identification of low-abundance phosphoproteins. *Analytical Chemistry* **74**, 607–616.

GYGI, S.P., ROCHON, Y., FRANZA, B.R. AND AEBERSOLD, R. (1999a). Correlation between protein and mRNA abundance in yeast. *Molecular and Cellular Biology* **19**, 1720–1730.

GYGI, S.P., RIST, B., GERBER, S.A., TURECEK, F., GELB, M.H. AND AEBERSOLD, R. (1999b). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology* **17**, 994–999.

GYGI, S.P., CORTHALS, G.L., ZHANG, Y., ROCHON, Y. AND AEBERSOLD, R. (2000). Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 9390–9395.

HANSEN, K.C., SCHMITT-ULMS, G., CHALKLEY, R.J., HIRSCH, J., BALDWIN, M.A. AND BURLINGAME, A.L. (2003). Mass spectrometric analysis of protein mixtures at low levels using cleavable 13C-isotope-coded affinity tag and multidimensional chromatography. *Molecular and Cellular Proteomics* **2**, 299–314.

HOCHLEITNER, E.O., KASTNER, B., FROHLICH, T. *ET AL.* (2005). Protein stoichiometry of a multiprotein complex, the human spliceosomal U1 small nuclear ribonucleoprotein: absolute quantification using isotope-coded tags and mass spectrometry. *Journal of Biological Chemistry* **280**, 2536–2542.

HUBER, M., BAHR, I., KRATZSCHMAR, J.R. *ET AL.* (2004). Comparison of proteomic and genomic analyses of the human breast cancer cell line T47D and the antiestrogen-resistant derivative T47D-r. *Molecular and Cellular Proteomics* **3**, 43–55.

IDEKER, T., THORSSON, V., RANISH, J.A. *ET AL.* (2001). Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934.

JAMES, P., QUADRONI, M., CARAFOLI, E. AND GONNET, G. (1993). Protein identification by mass profile fingerprinting. *Biochemical and Biophysical Research Communications* **195**, 58–64.

KEOUGH, T., LACEY, M.P. AND YOUNGQUIST, R.S. (2000). Derivatization procedures to facilitate *de novo* sequencing of lysine-terminated tryptic peptides using postsource decay matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* **14**, 2348–2356.

KLOSE, J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* **26**, 231–243.

KOLLER, B.H., GERAGHTY, D.E., DEMARS, R., DUVICK, L., RICH, S.S. AND ORR, H.T. (1989). Chromosomal organization of the human major histocompatibility complex class I gene family. *Journal of Experimental Medicine* **169**, 469–480.

LEMMEL, C., WEIK, S., EBERLE, U. *ET AL.* (2004). Differential quantitative analysis of MHC ligands by mass spectrometry using stable isotope labelling. *Nature Biotechnology* **22**, 450–454.

LILL, J. (2003). Proteomic tools for quantitation by mass spectrometry. *Mass Spectrometry Reviews* **22**, 182–194.

MOSELEY, M.A. (2001). Current trends in differential expression proteomics: isotopically coded tags. *Trends in Biotechnology* **19**, S10–S16.

MUNCHBACH, M., QUADRONI, M., MIOTTO, G. AND JAMES, P. (2000). Quantitation and facilitated *de novo* sequencing of proteins by isotopic N-terminal labelling of peptides with a fragmentation-directing moiety. *Analytical Chemistry* **72**, 4047–4057.

ODA, Y., HUANG, K., CROSS, F.R., COWBURN, D. AND CHAIT, B.T. (1999). Accurate quantitation of protein expression and site-specific phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 6591–6596.

ODA, Y., NAGASU, T. AND CHAIT, B.T. (2001). Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nature Biotechnology* **19**, 379–382.

O'FARRELL, P.H. (1975). High-resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry* **250**, 4007–4021.

ONG, S.E., BLAGOEV, B., KRATCHMAROVA, I. *ET AL.* (2002). Stable isotope labelling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular and Cellular Proteomics* **1**, 376–386.

QIU, Y., SOUSA, E.A., HEWICK, R.M. AND WANG, J.H. (2002). Acid-labile isotope-coded extractants: a class of reagents for quantitative mass spectrometric analysis of complex protein mixtures. *Analytical Chemistry* **74**, 4969–4979.

RAMMENSEE, H.G., FALK, K. AND ROTZSCHKE, O. (1993). Peptides naturally presented by MHC class I molecules. *Annual Review of Immunology* **11**, 213–244.

REYNOLDS, K.J., YAO, X. AND FENSELAU, C. (2002). Proteolytic 18O labelling for comparative proteomics: evaluation of endoprotease Glu-C as the catalytic agent. *Journal of Proteome Research* **1**, 27–33.

RINGROSE, J.H., MEIRING, H.D., SPEIJER, D. *ET AL.* (2004). Major histocompatibility complex class I peptide presentation after *Salmonella enterica serovar typhimurium* infection assessed via stable isotope tagging of the B27-presented peptide repertoire. *Infection and Immunity* **72**, 5097–5105.

SAPER, M.A., BJORKMAN, P.J. AND WILEY, D.C. (1991). Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *Journal of Molecular Biology* **219**, 277–319.

SCHENA, M., SHALON, D., DAVIS, R.W. AND BROWN, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470.

SCHMIDT, A., KELLERMANN, J. AND LOTTSPEICH, F. (2005). A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* **5**, 4–15.

SHEN, M., GUO, L., WALLACE, A. *ET AL.* (2003). Isolation and isotope labelling of cysteine- and methionine-containing tryptic peptides: application to the study of cell surface proteolysis. *Molecular and Cellular Proteomics* **2**, 315–324.

STEWART, I.I., THOMSON, T. AND FIGEYS, D. (2001). 18O labelling: a tool for proteomics. *Rapid Communications in Mass Spectrometry* **15**, 2456–2465.

STOCKLIN, R., ARRIGHI, J.F., HOANG-VAN, K. *ET AL.* (2000). Positive and negative labelling of human proinsulin, insulin, and C-peptide with stable isotopes. New tools for *in vivo* pharmacokinetic and metabolic studies. *Methods in Molecular Biology* **146**, 293–315.

TAO, W.A. AND AEBERSOLD, R. (2003). Advances in quantitative proteomics via stable isotope tagging and mass spectrometry. *Current Opinion in Biotechnology* **14**, 110–118.

UNLU, M., MORGAN, M.E. AND MINDEN, J.S. (1997). Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18**, 2071–2077.

WASHBURN, M.P., ULASZEK, R., DECIU, C., SCHIELTZ, D.M. AND YATES, J.R., III (2002). Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Analytical Chemistry* **74**, 1650–1657.

WEI, M.L. AND CRESSWELL, P. (1992). HLA-A2 molecules in an antigen-processing mutant cell contain signal sequence-derived peptides. *Nature* **356**, 443–446.

YALCIN, T., KHOUW, C., CSIZMADIA, I.G., PETERSON, M.R. AND HARRISON, A.G. (1995). Why are B ions stable species in peptide spectra? *Journal of The American Society for Mass Spectrometry* **6**, 1165–1174.

YAO, X., FREAS, A., RAMIREZ, J., DEMIREV, P.A. AND FENSELAU, C. (2001). Proteolytic 18O labelling for comparative proteomics: model studies with two serotypes of adenovirus. *Analytical Chemistry* **73**, 2836–2842.

YATES, J.R., III, SPEICHER, S., GRIFFIN, P.R. AND HUNKAPILLER, T. (1993). Peptide mass maps:

a highly informative approach to protein identification. *Analytical Biochemistry* **214**, 397–408.

YIN, Z., STEAD, D., SELWAY, L. *ET AL.* (2004). Proteomic response to amino acid starvation in *Candida albicans* and *Saccharomyces cerevisiae*. *Proteomics* **4**, 2425–2436.

ZHANG, R., SIOMA, C.S., THOMPSON, R.A., XIONG, L. AND REGNIER, F.E. (2002). Controlling deuterium isotope effects in comparative proteomics. *Analytical Chemistry* **74**, 3662–3669.

ZINKERNAGEL, R.M. AND DOHERTY, P.C. (1974). Restriction of *in vitro* T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature* **248**, 701–702.