Dynamics of the human nuclear proteome in response to DNA damage

Dynamics of the human nuclear proteome in response to DNA damage

Dynamiek van het humane celkernproteoom in reactie op DNA schade

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. W. H. Gispen, ingevolge het besluit van college voor promoties in het openbaar te verdedigen op maandag 11 september 2006 des middags te 4.15 uur.

door

Eef Hubert Cecil Dirksen

geboren op 17 juli 1978, te Eindhoven

Promotor
Prof. dr. A.J.R. Heck

Copromotor
Dr. M. Slijper

Financial support from
JE Jurriaanse Stichting
GE Healthcare (Amersham Biosciences)
is gratefully acknowledged

Aan Marchien

ISBN
ISBN-10: 90-393-4297-0
ISBN-13: 978-90-393-4297-8

Cover
Photography by Marco de Groot

Circumstances on the bus platform in the morning closely resemble some of the early dynamics of the human nuclear proteome in response to DNA damage: upon emergence of a damaged DNA strand (bus)*, apparently unordered, but watchful, repair proteins* (people) *approach the strand and form distinct foci near the damage sites* (bus doors)*. In the cell, these foci are DNA repair hotspots.*

Printing
Ponsen & Looijen BV, Wageningen

Contents

Chapter 1

General Introduction

I. DNA damage, repair and carcinogenesis

All cells undergo division cycles throughout their life span. During the process of cell division, everything within the cell must be duplicated in order to ensure the survival of the resulting daughter cell. Accurate, efficient and rapid duplication of the cellular genome, which holds all genetic information necessary for cell functioning, is of particular importance for cell survival. The integrity of the genome however, is constantly challenged by both endogenous (1) and exogenous factors, such as ionizing radiation (IR, for example X-rays during diagnostic radiology) or compounds in food and beverages, such as acrylamide and alcohol that can induce a variety of types of damage (2). Of these, the most severe is the DNA double strand break (DSB), which is generated when two complementary strands are broken simultaneously at sites that are close enough so that both base pairing and chromatin structure are not sufficient to keep the strands juxtaposed. This can result in physical separation of the two strands, whereby bases at the newly formed DNA ends often have additional damage as a result of oxidative side reactions. This poses the cell with a major challenge in repairing this damage.

Even though DSBs form a major threat to genome integrity, their formation also occurs naturally during meiosis I (3) and in B- and T-lymphocytes during V(D)J recombination, which provides the basis for the antigen-binding diversity of immunoglobulins (4). Therefore, the cell has a number of mechanisms to sense and repair DSBs that collectively make up the DNA damage response (2,5-9).

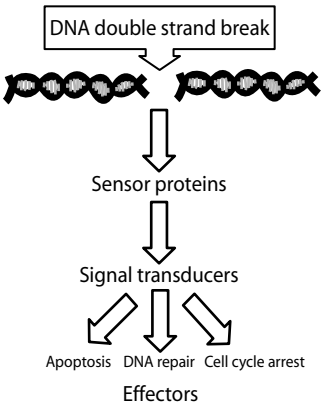


Figure 1. *Schematic overview of the DNA damage response. Upon detection of a lesion in DNA by sensor proteins, transducers amplify and diversify these signals and in turn activate a number of effector proteins involved in for example cell cycle control and DNA repair.*

As illustrated in Figure 1, the DNA damage response consists of three consecutive levels: the so called *sensor* proteins that act at the first level (Ku70/80, (10,11), RPA (12,13)) recognize lesions in DNA (5,6,14,15). These proteins in turn activate a *transducer* system which is a signal transduction cascade that makes up the second level of the response (ATM / ATR (16-19) DNA-PK (17,20,21)). Damage signals are both amplified and diversified and a number of *effector* pathways (the third level) are triggered.

Important effector pathways are those controlling the cell cycle at the checkpoints. In order to prevent damaged DNA from being replicated, the cell can activate the G_1/S checkpoint, which results in cell cycle arrest and prevents cells from entering S phase by blocking replication (22-24). When DNA damage is induced during S phase, or when damaged DNA escapes the G_1/S checkpoint, the intra S phase checkpoint can be set off, leading to a block in replication (25). A third checkpoint at the end of the G_2 phase prevents cells from undergoing mitosis with damaged DNA present (15). Subsequently, other effector pathways can be triggered, leading for example to the repair of damaged DNA (26) or the onset of apoptosis –programmed cell death- in case of severe DNA damage (27).

DNA repair pathways

There are two important pathways of DSB repair (14): homologous recombination (HR (28-31)) and non-homologous end joining (NHEJ (32)). These pathways are largely distinct but function in complementary ways. In the first pathway, sequence information from an undamaged DNA molecule with which the damaged strand shares extensive homology is used to repair the damaged chromosome. This leads to accurate repair of DNA, while ligation without the requirement of extensive homology between the individual strands occurs in the second pathway. Consequently, this repair is less accurate. Even though proteins from both pathways are highly conserved in evolution, HR is mainly used for DSB repair in prokaryotes and lower eukaryotes. In mammals, NHEJ predominates particularly in the G_0 and G_1 phase of the cell cycle and HR is particularly important during G_2 and S phase. Figure 2 gives an overview of the stages that occur in both repair pathways.

In NHEJ, Ku70 and Ku80 play an important role in DNA DSB recognition. The Ku proteins form a heterodimer that binds DNA surrounding a DSB in a non-sequence-dependent fashion, and acts as the recruiting subunit for the DNA dependent protein kinase (DNA-PK). Upon binding, DNA-PK displays protein serine and threonine kinase activity. *In vivo* substrates for this kinase include XRCC4 (33) and replication protein A2 (12) that, when phosphorylated, facilitate NHEJ.

Another factor in mammalian NHEJ is DNA ligase IV that forms a complex with XRCC4 and brings about DNA strand-joining events. HR on the other hand, follows a more complex mechanism that has only been partially elucidated using studies in bacteria and yeast. Since we are primarily interested in NHEJ, HR is not further explained here.

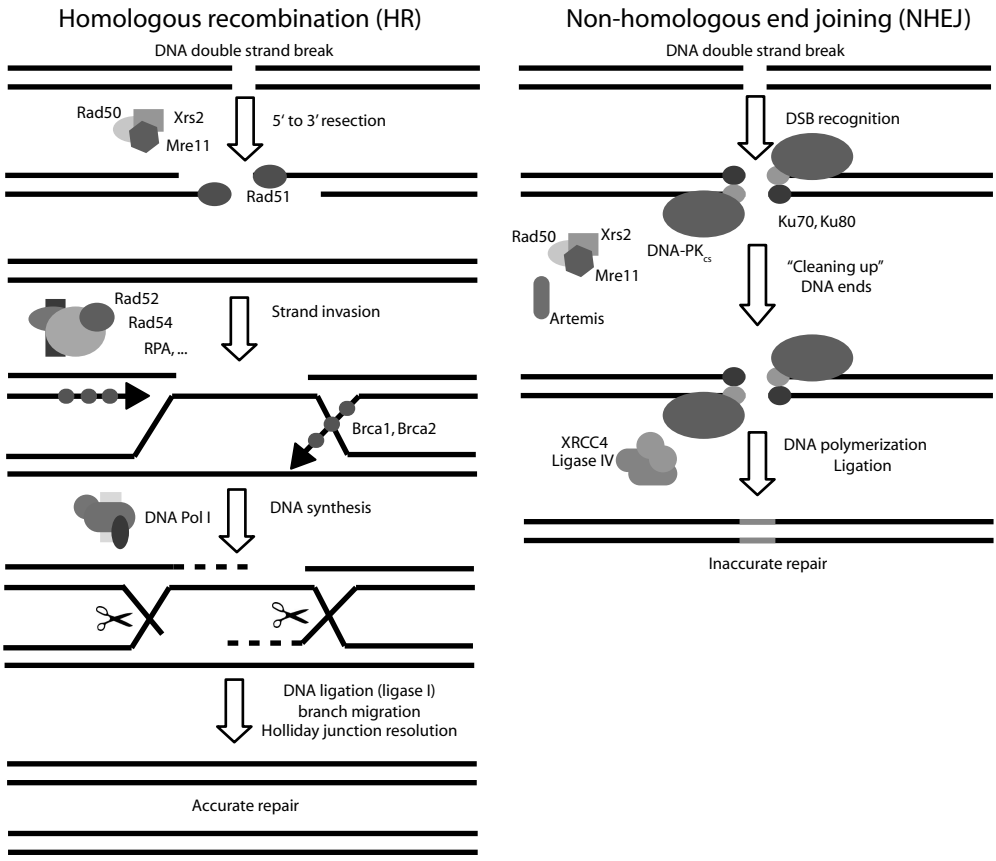


Figure 2. *General outline of the pathways for DNA double strand break repair (14). In mammalia, HR involves factors from the Rad50 group, Mre11 and NBS1 proteins. It starts with nucleolytic resection of the DSB in the 3' 5' direction. Subsequently, the ensuing 3' tails are bound to Rad51, Rad52 and Rad54. This filament then interacts with an undamaged DNA molecule. DNA polymerase extends the 3' terminus of the damaged DNA molecule. It copies information from the undamaged molecule and the ends are ligated by DNA ligase I. After that the DNA cross-overs (Holliday junctions) are cleaved, yielding two intact DNA molecules (28). Events in NHEJ are discussed in the text.*

DNA repair in a chromatin environment

As DNA is densely packed in chromatin, DSBs are not always easily accessible. Therefore, chromatin decondensation has to take place prior to DNA repair in a similar way as occurring prior to transcription (34). Over the past years, research has been increasingly aimed at chromatin remodelling and the influence of core histone variants and modifications in particular. It is clear that very specific modifications occur upon induction of DNA double strand breaks on particular histones, such as phosphorylation of human histone H2A.X on S^{139} by members of the PI3K-like family of protein kinases (35), which include ATM, ATR and DNA-PK. This modification has been described to be a beacon for other protein (complexes) involved in DNA repair (36,37).

The current knowledge however resembles only the tip of the iceberg, and recently new modifications, which influence processes in the DNA damage response have been detected, such as core histone acetylation (38), methylation (39,40) and ubiquitination (for reviews see (41,42)). The unravelling of the so-called histone code, which programs transcription either through silencing of genes or through increased transcription by way of chromatin decondensation, has been extended to a possible epigenetic code for the repair of DNA DSBs (43). Next to variants and modifications, histone chaperones, which are proteins characterized by the presence of acidic (glutamate- and aspartate-rich) regions also play an important role in the structural organization of chromatin. These proteins disrupt DNA-histone contacts and thereby aid in DNA-related processes. Several parallels have been described between chromatin remodelling that occurs in transcription regulating processes and DNA repair, in which the same proteins act (38).

DNA damage and carcinogenesis

Interactions between proteins in the DNA damage response pathway are tightly regulated. Therefore, a gene mutation leading to, for example, an impaired protein that is important for DNA repair can have a dramatic effect on downstream protein functioning. A well-known example is p53, which is a key protein in the DNA damage response (44). This protein regulates transcriptional control of target genes involved in several stress responses, including that to DNA double strand breaks (45). Expression of mutant forms of p53 is known to alter cellular resistance to DNA double strand breaks (46) and since p53 is involved in cell cycle checkpoint activation and apoptosis, impaired functioning also affects cell proliferation. On top of this, p53 is subject to a variety of post-translational modifications (47,48), such as phosphorylation (49-52), acetylation (53-56), ubiquitination (55,57) that can either influence its turnover rate, cellular

localization and/or its functioning. Hence, p53 functioning can also be directly impaired when the enzymes responsible for any of these post-translational modifications do not function properly. Consequently, the p53 tumour suppressor protein is involved in the onset and development of several types of cancer, such as hepatocellular carcinoma, colorectal, lung, bladder and breast cancer, brain tumours, squamous cell skin carcinoma, leukaemia and lymphoma [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM].

Another well-known example of loss of function mutations occur in the *ataxia telangiectasia* mutated kinase that phosphorylates a number of substrates in response to DNA double strand breaks (16,19,58-60). In *ataxia telangiectasia* (AT), this enzyme is present in cells, but several mutations in the amino acid sequence (61,62) severely affect its functioning. This leads to a seriously impaired DNA damage response and increases cellular sensitivity towards ionizing radiation, genomic aberrations and malignancies (63).

Even though the molecular mechanisms underlying the increased sensitivity towards ionizing radiation seen in AT have been extensively studied over the past decades (63-66), several studies have aimed at identifying genes involved in the DNA damage response *in general* (67,68). This type of research can help to further explain this cellular response or to elucidate other syndromes with similar phenotypes. Another example in which an increased sensitivity towards DNA damage-inducing agents is observed, is head and neck squamous cell carcinoma (HNSCC). It is known that besides well-known risk factors like smoking and alcohol consumption, an intrinsic susceptibility plays a role in HNSCC carcinogenesis (69). This susceptibility is reflected by the mutagen sensitivity, which is determined by the number of chromatid breaks per cell (b/c) induced after exposing patient lymphocyte cells to a DNA damaging agent (70). A large-scale study involving cancer patients, their healthy family members (including homozygotic and heterozygotic twins) as well as family members suffering from HNSCC, showed that mutagen sensitivity is an inheritable factor (71) and a b/c value dichotomised at 1.0 was found to be the best predictor of a hypersensitive phenotype (69). An increased number of b/c (larger than 1.0) correlates with decreased DNA stability, which is a well-known risk factor for cancer development (72). Patients suffering from HNSCC that developed a single primary tumour (SPT) were found to show a significantly higher sensitivity towards chemically induced DNA damage compared to control subjects as determined in the mutagen sensitivity assay. Consecutively, HNSCC patients that developed multiple primary tumours showed a significantly higher sensitivity compared to SPT patients (73). Even though mutagen sensitivity has become a potential biomarker for the susceptibility to HNSCC development (74), the challenge

remains to track genes involved in the response to DNA damage in general and in increased susceptibility of certain persons to DNA damage.

In this thesis, I will focus on the biochemical mechanisms underlying the development of cancer and the role of double strand DNA breaks in this process. Bearing in mind that the response to DNA damage is a constitutional factor, it can be studied in different types of cells, such as lymphoblast cells. Lymphoblast cells are often used as a model system to study the response to DNA damage (75-77). After extraction of lymphoblast cells from blood of individuals, the cells can be immortalized through a transformation with the Eppstein-Barr virus, as described by Hsu *et al*. in 1990 (78). By immortalizing cell lines, their lifetime is extended, since they can replicate indefinitely. This introduces flexibility in large-scale, long-term studies, because several experiments can be performed multiple times using the same cell line(s), which in turn increases the confidence of the results obtained. This also introduces the possibility of performing genetic and proteomic analyses of the same cells and allows strict control of the experimental conditions. Hsu *et al*. also studied the effect of immortalisation on the mutagen sensitivity phenotype of the lymphoblast cell lines used and found that it was similar to that of the lymphocytes short-term cultures (78), which meant that this cell line could be used for studies of the DNA damage response. In our study, lymphoblastoid cell lines were used that were derived from blood drawn of individuals that were involved in the large-scale study by Cloos *et al*. (71). The phenotype of the cells, i.e. susceptibility towards DNA double strand breaks, was also checked here and was found to be unchanged after immortalisation.

When focussing on the effect of double strand DNA breaks on cellular functioning, ionizing radiation (IR) can be used to induce DNA damage. This procedure however has some practical drawbacks, resulting in for example inhomogeneous irradiation. A more convenient way to induce this specific type of DNA damage is through the use of bleomycin, a radiomimetic compound that induces DNA double strand breaks (79). This compound can be added to the growth medium, is taken up by cells and subsequently induces DNA damage in a relatively short period of time. Bleomycin requires molecular oxygen and an activating metal, such as iron or copper in order to be active (80). An additional advantage of bleomycin over ionizing radiation is the fact that since both agents also induce single strand DNA breaks, the ratio DSB/SSB for bleomycin is higher compared to that for IR (1/9 *vs* 1/100, respectively) (81). Cells will respond to the induced damage by activating DNA damage response pathways in order to start DNA repair. Next to that, cells will have to deal with secondary effects,

such as the cellular stress that is induced by the oxidizing action of bleomycin. This is similar to the secondary effects of ionizing radiation that are caused by radicals that are formed during oxidative DNA cleavage.

***II**. Techniques to study the overall cellular response to DNA damage*
Transcriptomics

Upon induction of DNA damage through exposure to IR or radiomimetic compounds, the cellular response to DNA DSBs affects several processes ranging from transcription and translation to protein post-translational modifications. This can be studied at different levels, as illustrated in Figure 3. First of all, micro-arrays can be used to get an overview of the response of large sets of genes to this particular stimulus (82) in a set-up that is referred to as a (functional) genomic or transcriptomic approach.

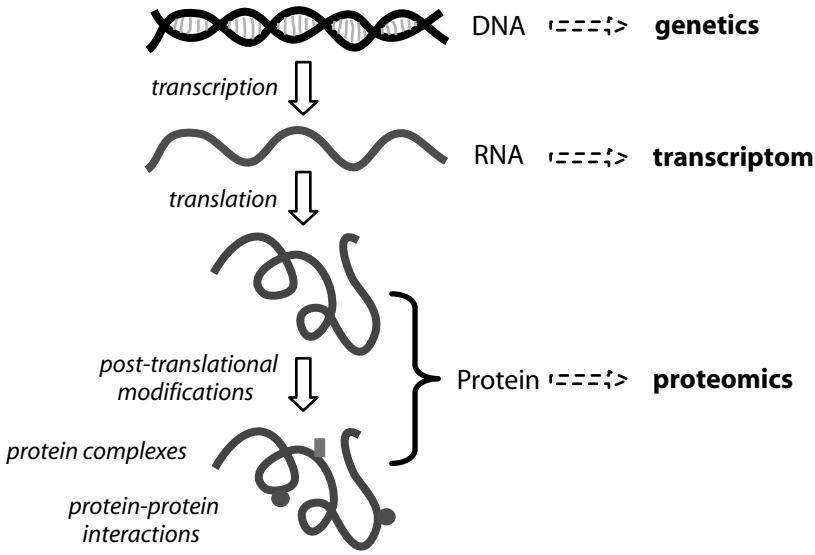


Figure 3. *Conversion of genetic information into functional proteins by subsequent transcription, translation and protein-protein interactions. The complexity of the resulting products increases due to alternative splicing and post-translational modifications. The levels at which these processes can be studied are given on the right side.*

In a transcriptomic experiment mRNA levels are determined, which is a measure for changes in gene expression as a result of a particular stimulus, like for example DNA damage induction. Amundson, *et al* have carried out a number

of transcriptomic experiments to determine the response of human cell lines to ionizing radiation and DNA damaging chemical compounds (83-87). In one of these studies potential biomarkers of ionizing radiation were identified in *ex vivo* irradiated human peripheral white blood cells. Resulting candidate genes such as CDK1N, DDB2 and GADD45A were validated *in vivo* in patients undergoing total body irradiation using RNA extracted from whole blood.

A transcriptomic approach was also used to assess mRNA levels of genes in lymphoblastoid cell lines at various time points within 24 hours following ionizing radiation. It was found that genes involved in cell cycle control, DNA repair, DNA metabolism, RNA processing and cell death were differentially regulated. For a number of other differentially regulated genes no direct relation to the DNA damage response could be established (77). These proteins are, for example, involved in detoxification pathways or in maintaining the redox balance of the cell, which is disturbed upon induction of cells with ionizing radiation or DNA damaging chemicals.

Although transcriptomics can provide insight into the regulation of genes upon certain stimuli, the actual cellular response is effectuated by proteins. Moreover, the correlation between gene and corresponding protein expression regulation in mammalian cells has been reported to be poor (88,89). This is mainly caused by differences in turn over rates of mRNA and corresponding proteins and by the influence of protein post-translational modification. The latter is particularly the case in the DNA damage response in which protein activity is controlled by a variety of kinases and phosphatases. This means that proteins already present in the cell can be activated or inactivated by post-translational modifications, resulting in a cellular response largely independent of gene regulation. Consequently, a second approach to study the effect of certain stimuli that reflects what occurs at the cellular level, is to focus on changes in expression levels of all proteins present in a cell between unstimulated cells and cells in which DNA damage was induced. This complement of all proteins present in a cell at a certain point in time is called the proteome and the study of that entity is called proteomics (90) (see Figure 3).

2D gel-based proteomics

In proteomics research, all proteins in complex mixtures are analysed simultaneously. This requires methods that can resolve individual proteins in a particular sample with high resolution and in a reproducible way. Two-dimensional gel electrophoresis (2D-GE), a method frequently used to accomplish this, was already introduced in 1975 by O'Farrell (91) and is based on separating proteins using two of their biophysical properties. The first is the pI that is used

to separate proteins by iso-electric focusing (IEF) in the first dimension. The pI, the pH at which the net charge of a protein is 0, is determined by the contributions of all positively and negatively charged amino acids. Secondly, proteins are separated according to their molecular weight by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) in the second dimension.

In practice, proteins are taken up into a polyacrylamide gel in which a pH gradient is created by modification of acrylamide with compounds containing either carboxylic or amino groups. Upon application of an electrical field, proteins migrate until they reach their pI at which the net charge of the protein is 0. After application of the IPG strip with the focussed proteins onto the polyacrylamide gel for the second dimension, SDS in the running buffer surrounds proteins with a negatively charged shell that is proportional to their size/weight. This separates proteins in the gel while migrating from the anode to the cathode: small proteins move about quickly and end up in the lower side of the gel, while large proteins travel much slower and therefore end up in the upper part of the gel. There is a linear relationship between the logarithm of the molecular weight and the relative migration distance of a protein.

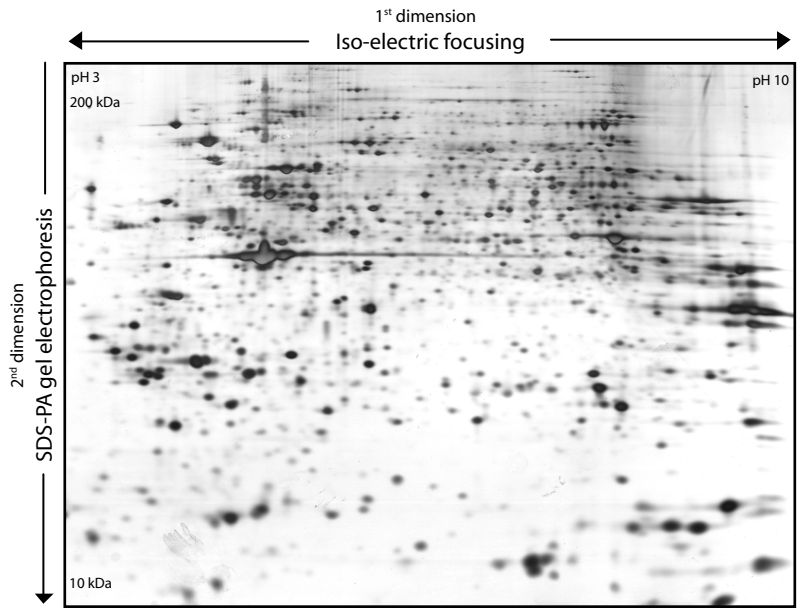


Figure 4. *Typical example of a silver stained two-dimensional gel. Nuclear lysate (150 µg) was separated in the first dimension on a 24 cm non-linear pH 3-10 IPG strip and subsequently in the second dimension on a continuous 12.5% SDS-polyacrylamide gel. Single spots represent individually resolved proteins.*

In order to allow the comparison of protein expression levels between multiple samples, proteins need to be visualised. This is routinely done by using either Coomassie blue or silver staining (92,93). After staining, the intensity of a protein spot correlates to the protein quantity in the sample and the resulting 2D gel can be considered as the protein equivalent of a micro-array, the major differences being that the position of the proteins within the gel is not regular as the spotted cDNA clones on a micro-array. Moreover, while the identity of spotted sequences on a micro-array is known, that of a protein at a certain position in the gel has to be determined afterwards (See Figure 4).

Examples of 2D-GE based proteomic approaches applied to the DNA damage response are studies by Szkanderova *et al* (94-97) in which the effect of γ-radiation on L929 and T-lymphocyte leukaemia cells is described. Several metabolic enzymes and proteins involved in regulation of the cellular redox balance that were found to be regulated were identified. Another study describes the effect of UV light radiation on the proteome of HeLa cells (98), in which several actin isoforms and heat shock proteins were found to be differentially regulated.

The results mentioned above are obtained from experiments performed with single cell lines. This compromises the confidence of the proteins that are found to be differentially regulated as a result of the biological phenomenon studied, since the response of particular cell lines to stimuli might differ greatly. As a result, changes in protein expression levels independent of the DNA damage response can appear as false positives. In general, basal protein expression levels show differences, particularly when using human cell lines. This is the reason why lists of 'differentially regulated' proteins often contain similar proteins independent of the stimulus. To avoid this, the correct design of an experiment is of utmost importance (99). This comprises acquiring sufficient biological replicates, optimisation of sample preparation, such as the purification of a particular organelle, and the experimental set up. In the case of 2D-GE, the experimental set-up determines the size of the pH gradient and the subsequent gel, the number of analytical replicates, but also methods for gel visualization and subsequent statistical analysis.

Despite its large resolving power, 2D GE has some well-known disadvantages: the procedure of running a 2D gel is laborious and thus time consuming, it is hard to automate and suffers therefore from low reproducibility. Next to that, physical characteristics of particular subgroups of proteins, such as hydrophobicity and extreme molecular weights or pI values, generally cause

underrepresentation in a 2D gel. Ultimately, the general methods for protein visualization, like Coomassie blue and silver staining, limit protein quantitation capacities of proteins separated on a 2D gel due to poor sensitivity and a limited linear dynamic range (100-103).

2D difference in-gel electrophoresis

The shift in focus of proteomics research from qualitative detection of proteins in gel to quantitative analysis of protein expression levels between samples (gels), has led to the development of new ways of protein visualisation, based on covalent or non-covalent labeling of proteins with fluorescent dyes. After the introduction of the SYPRO dyes (104), that are used to visualise proteins after they have been in-gel separated, fluorescent dyes were developed that covalently label proteins *prior* to in-gel separation (105-107). The labeling is based on modification of the ε-amino group of lysine with an NHS-activated fluorophore molecule and to minimize the influence of labeling on protein migration, the fluorescent labels have been size and charge matched. Due to the positive charge present on the label itself, which replaces the positive charge on the lysine side chain that is lost upon labeling, the pI will not differ between a labelled and an unlabeled protein, rendering the pH at which the protein focuses unchanged in the first dimension. The influence on migration in the second dimension will also be minor, since the mass of a labelled protein will only increase by 450 Da compared to its unlabeled equivalent. Initially, two spectrally resolvable labels –Cy3 and Cy5, originating from DNA micro-array analyses – were used, allowing the simultaneous analysis of two protein samples in the same gel. Recently, a third label was introduced which, instead of being used for the labeling and analysis of a third analyte sample, is routinely used to label an internal standard that consists of equal amounts of all the samples in the experiment. The advantage of using an internal standard is twofold: matching confidence between gels increases and it enables separation of technical variation, which still is one of the major drawbacks of 2D-GE, from sample/biological variation (108).

As calculations on dynamic protein expression levels are based on comparisons between spot intensities of the same protein on multiple gels, reproducibility is crucial in proteomics approaches. By using three different, spectrally resolvable fluorescent dyes, three individual samples can be labelled and subsequently separated on a single gel.

Depending on the amount of protein available for analysis, two different approaches can be chosen. The first is minimal labeling, in which about 3% of the proteins is labelled on the ε-amino group of lysines. The unlabeled protein

can then be used for protein identification and characterization. If the sample amount is limited however, a second approach called saturation labeling can be applied. This involves complete labeling of proteins on cysteine residues and has the advantage that fluorescence signals are maximized for the proteins in the mixture. A disadvantage is that the analysis is limited to proteins containing cysteine residues. The chance of finding a protein that contains no cysteines is more than 3 times higher than the chance of finding a protein lacking lysines.

The individual images from every gel are acquired by measuring the emission after fluorophore-specific excitation. Since the same proteins that are present in all three samples have migrated to the exact same positions in the gel, their fluorescence can be directly compared and provides information about differences in protein levels between two samples.

The increasingly complex set-ups of 2D gel-based proteomic experiments require dedicated image analysis software packages that can simultaneously analyse large numbers of spots within multiple gels. Key part of the image analysis is detecting and matching of protein spots across all gels in the analysis set, to ensure that the intensities of the same protein spots are compared. Small differences induced by, for example, air bubbles, inhomogeneous polymerisation of acrylamide, streaking or smearing, can greatly influence the pattern of a 2D gel. Therefore spot detection and matching algorithms have been designed by a variety of manufacturers (for a review on image analysis software packages, see (109)), such as PDQuest (BioRad), Progenesis (Nonlinear Dynamics), Melanie (GeneBio) and DeCyder (Amersham Biosciences). The inter-gel matching algorithm of DeCyder, for example, is based on pattern recognition that matches an individual spot in one gel with an individual spot in another gel based on its neighbouring spots. By manually defining matched spots ('landmarks') the accuracy of this procedure is improved and when a substantial part of the number of detected spots has been matched to the majority of the images that is included in the analysis, spot intensities can be compared.

To make certain that changes in protein expression levels are calculated as a function of the levels present in untreated or unstimulated cells, all images have to be normalized. In DeCyder, normalization of fluorescence intensities occurs at two levels: the differential in-gel analysis (DIA) normalization works on the individual gel images. For this, the spot ratios are calculated after background subtraction and these are fitted into a data histogram, assuming that the majority of all protein spot intensities in the experiment do not change between two situations, to ensure that the data fit a normal distribution. After optimisation of the curve using a least means square gradient descent algorithm,

the histogram is recalculated and the standard deviation is determined. In the biological variation analysis (BVA) normalization the data histogram is optimised and after that, the centre of the curve is denoted as the centre of volume, which is subsequently used to calculate normalized spot volumes. The normalization methods mentioned above are embedded in the image analysis software. Additionally, some research groups have developed independent methods for the analysis of data obtained from 2D-DiGE experiments (110) that provide a better insight in the calculation procedures. Finally, statistical analyses such as the Student's T-test or an analysis of variance (ANOVA), give a numerical level of confidence to determine whether a particular change in protein abundance is above biological variation (111). In order to study large (2D-DiGE) datasets in more detail, additional statistical tests have been developed. Many of these originate from micro-array analysis experiments and have been adapted for application on 2D gel data (112).

LC-based proteomics

In an attempt to circumvent the aforementioned problems related to two-dimensional gel-based proteomic approaches, liquid chromatography-based methods have been developed for the separation of complex biological samples. One of the main differences between gel-based and LC-based approaches is that in the latter proteins from a cellular lysate are proteolytically digested prior to separation and analysis.

Typically, a digested lysate obtained from cells in state A is separated and analysed using reversed phase liquid chromatography (RP-LC) coupled to mass spectrometry. In such an experiment, peptides are loaded on a stationary phase (C_{18}) on which they absorb under aqueous conditions as a result of their hydrophobic properties. Since peptides show differences in hydrophobicity, they can be gradually eluted from the stationary phase using an increasing amount of organic modifier in the mobile phase changing the conditions from aqueous to organic. The retention times of all peptides in this sample can subsequently be compared to those from the lysate of cells in state B. This however, introduces similar problems as encountered in standard 2D-GE concerning reproducibility of retention times and internal standardization of ion intensities, which compromises accurate quantitative comparisons. To further improve this, stable isotope labels have been introduced in gel-free quantitative proteomic experiments (for a review see (113)). In general, stable isotopes have the advantages that they are physically the same as naturally occurring isotopes apart from the fact that their masses differ. Incorporation of stable isotopes into proteins and/or peptides can be used to directly compare cellular states, since

the labeled and the unlabeled sample can be mixed and subsequently analysed in the same LC-MS experiment. Finally, the mixed sample is analyzed in the mass spectrometer that can discriminate between the light and heavy variant of the stable isotope labels used. This is highly analogous to the fluorescent approach, in which differently labeled samples are mixed and finally the protein levels of each sample are detected individually again.

Various strategies for the incorporation of stable isotopes have been described over the past years, as can be seen in Figure 5, ranging from metabolic stable isotope labeling (114) and stable isotope labeling of amino acids in cell culture (SILAC (115,116)), which are the methods that enable the earliest introduction of labels into proteins and therefore provide the most accurate way of protein quantitation, to ICAT labeling (117) and tryptic digestion in $H_2{}^{18}O$ (118).

In the latter two methods, as well as in iTRAQ (119) and other peptide derivatization methods, the stable isotope label is only introduced after several steps of sample preparation, such as cell lysis and/or proteolytic digestion. As a consequence, quantitative information is partly lost as a result differences in losses that can occur during those sample preparation steps, which compromises accuracy.

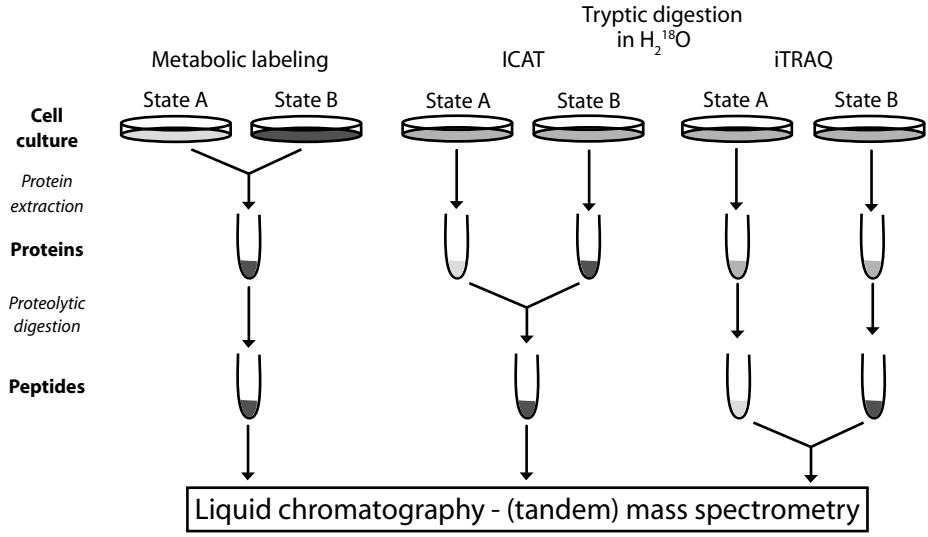


Figure 5. *A choice of strategies for the labeling of proteins and peptides with stable isotopes for accurate quantitation purposes.*

At present, it is clear that even though multidimensional LC approaches were originally meant to replace 2D GE, the two techniques are in fact complementary. Illustrative examples are hydrophobic membrane proteins on the one hand, which are hard to analyze using 2D GE due to their poor solubility, and complex protein isoforms, resulting from alternative splicing, single nucleotide polymorphisms and post-translational modifications on the other. The latter are preferably analyzed using 2D GE because of the fact that this technique is able to resolve protein isoforms, at least when they have different pI's and/or molecular weights.

***III**. Mass spectrometry in proteomics*

Following statistical analysis of a set of 2D gels, a subset of proteins is typically taken of which the abundance changes between the different biological samples analysed. These proteins are selected for spot excision from a 2D gel and subsequent proteolytic digestion. This is routinely done using trypsin (120,121) that degrades proteins by hydrolysing the C-terminal peptide bond after a lysine or an arginine residue. The resulting peptide mixture can be used for identification of the protein (122-124). The latter can be achieved using mass spectrometry that has been playing an increasingly important role in proteomics since the introduction of two ionisation techniques, matrix assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI). These enable the analysis of large biomolecules, such as peptides and proteins.

Matrix assisted laser desorption/ionisation

In MALDI (125), the analyte is mixed with an acidic matrix solution and applied on a target plate. During drying of the sample, the matrix crystallizes thereby enclosing the peptides. These crystals, containing sample and excess matrix are bombarded with laser light. The laser energy applied is absorbed by the aromatic matrix molecules and subsequently transformed into excitation energy that causes desorption of matrix and analyte molecules. Although the majority of the desorbed species are neutrals, positive and negative ions are formed in the plume as well. Numerous collisions occur in this desorbed matter that may ultimately lead to ionisation of peptides by a proton transfer reaction that generates a set of singly charged ions representative for the peptides in the mixture. Since peptides are indirectly ionised after energy transfer by the matrix molecules, MALDI is considered to be a 'soft' ionisation method, making it better suited for the analysis larger (labile) biomolecules compared to 'old-fashioned' ionisation techniques such as electron impact (EI) or fast atom bombardment (FAB).

The pulse of generated ions is normally analysed using a time-of-flight (TOF) mass analyser (126), which is sensitive, relatively simple and applicable over a high mass range. All ions are accelerated with the same kinetic energy (U_k) into the flight tube. Subsequently, the time individual ions need to reach the detector at the end of the flight tube is measured. As the kinetic energy equals half the mass multiplied by their velocity squared ($U_k = ½mv^2$), the velocity of the ions is inversely proportional to the square root of their masses. As a result of this, ions with different masses have different flight times. After determining the flight times of all ions through the flight tube, a mass spectrum is acquired. To improve the relatively low resolution of the linear TOF analyser, which is a disadvantage of the linear TOF, a reflectron can be used that basically lengthens the flight path and consequently increases the resolving power by reducing the spread in kinetic energy of the ions. Additionally, time lag focussing has helped to further improve resolution and mass accuracy that can be obtained using MALDI-TOF mass spectrometry (127).

Electrospray ionisation

In ESI (128) (reviewed in (129,130)) the analyte (peptide mixture) is usually dissolved in a solution containing organic solvent and a volatile acid. This solution is forced through a needle to which a potential is applied. This process is referred to as pneumatic nebulization and produces large charged droplets. The voltage applied (1-5 kV) disperses the emerging solution into a fine spray of droplets. Subsequently, the solvent evaporates, decreasing the droplet size and increasing the charge density on the droplet's surface. When Coulombic repulsion overcomes the droplet's surface tension, droplet fission occurs. This produces series of smaller and lower charged droplets and eventually leads to the formation of desorbed analyte ions. In positive ion ESI-MS, peptides become charged as a result of multiple protonations, preferably at amino groups at the N-terminus of a peptide or arginine, histidine and lysine side chains. ESI consequently generates multiply charged ions, which provides a large advantage since it allows analysis of large biomolecules on an instrument that has a relatively low *m/z* range. The introduction of nanoflow ESI (120,131) has played an important role in the development of mass spectrometry for proteome analysis. This miniaturized version of ESI that requires only a few nanoliters per minute allows very sensitive detection of peptides and proteins. Since in ESI analytes are directly ionised out of a solution, it is routinely coupled to liquid chromatography based separation methods (132). The ions generated by ESI can also be analysed using a TOF detector, but more frequently other mass analysers

are used for this purpose, such as quadrupole TOF (133) or ion trap (linear ion trap (134) or quadrupole ion trap (135)) instruments.

The ionisation methods discussed above have been implemented in proteome research and are mainly used to identify proteins from a mixture of peptides generated from a 2D gel-separated spot, a digested 1D SDS-PA gel lane, or from an in solution digested mixture of proteins. Two mass spectrometric approaches that can be used to identify proteins are peptide mass fingerprinting and peptide sequencing. Despite the wide application of both of these methods, peptide sequencing is used more and more at the cost of peptide mass fingerprinting.

Peptide mass fingerprinting

A mixture of peptides can be analysed using matrix assisted laser desorption/ionisation-time of flight mass spectrometry (MALDI-TOF MS) (136).

Since proteins on the one hand are built up of 20 building blocks, amino acids, but on the other hand have different amino acid sequences, peptide fragments resulting from proteolytic processing are unique for every protein and have unique masses. It is therefore referred to as a peptide mass fingerprint (137-140). After acquiring the fingerprint and conversion of the data into a text file containing the individual peptide masses and their relative intensities, the masses are compared to theoretical masses present in a database, such as the SwissProt database [http://www.expasy.org/sprot], using an interface like Mascot [http://matrixscience.com/search_form_select.html].

Additional information is required to get a successfull identification, such as the taxonomy of the species studied, the proteolytic enzyme used, whether any fixed (carbamidomethylation of Cys) or variable (oxidation of Met) modifications can be expected and the peptide tolerance (usually 50 ppm). In turn, a (set of) protein(s) that suits the given criteria best will be returned. Confident protein identification does not require that all peptides originating from a single protein are retrieved: a sequence coverage –which is the percentage of the total protein sequence that can be read from the peptide mass fingerprint- of about 10-20% is usually enough.

Peptide sequencing

In case peptide mass fingerprinting has not led to unambiguous identification of a digested protein, a peptide mixture can also be analysed using tandem mass spectrometry. For this, the generated ions are subjected to collision-induced dissociation (CID), a process in which the selected peptide ion

collides with inert gas molecules, like argon. Upon this collisional activation, the added internal energy leads to the breaking of peptide bonds and fragment ions are formed. For particular series of fragment ions, such as *b* and *y* ions, the difference in mass corresponds to the residue mass of one of the twenty amino acids. This can be used to read the peptide sequence from either the N-terminal or the C-terminal side. Next to that, other types of ions, such as immonium ions, also provide information about the composition of the peptide. An overview the nomenclature for fragment ions (141) is given in Figure 6.

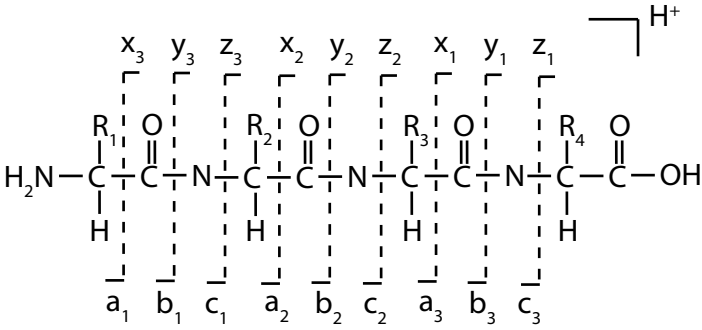


Figure 6. *Roepstorff-Folmann nomenclature of peptide fragmentation (141).*

With this information, part of the amino acid sequence of a selected peptide is determined and subsequently, the protein from which it originated can be identified. Again, due to the fact that proteins are built up of 20 amino acids, a sequence tag, which is a partial amino acid sequence is usually enough to uniquely identify the protein. The length of this tag, required for a protein identification, is dependent on the size of the genome that is studied: for bacteria a tag of 3 to 4 amino acids can be enough, while 6-8 residues are required in the case of human proteins.

***IV**. Analysis of protein phosphorylation*

As mentioned in the beginning of part II, post-translational modification is an important parameter that influences the activity of a protein (see also Figure 3), and consequently cellular processes.

One of the most abundant post-translational modifications is a covalent modification by a protein kinase of hydroxyl groups of serine, threonine and tyrosine with phosphate (142). It is hypothesized that one third of all proteins in

a cell are phosphorylated at a given moment and the ratio of serine/threonine/tyrosine phosphorylation is estimated to be 90:10:0.05 (143). Although I recognize the importance of tyrosine phosphorylation (144) especially in receptor-mediated signalling, I will focus on the analysis of serine and threonine phosphorylation only.

In turn, phosphatases can dephosphorylate phosphorylated amino acids, see Figure 7. This kind of on/off switch partly determines the activity of a protein within a signal transduction cascade.

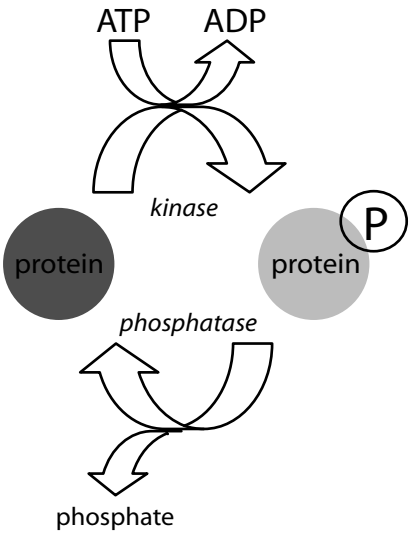


Figure 7. *Schematic representation of the dynamic interplay between kinases and phosphatases that regulates protein activity.*

The study of the subgroup of cellular proteins that carry this modification is often referred to as phosphoproteomics. Mass spectrometric strategies used in this discipline frequently yield dissatisfying results, which is generally considered to be caused by physical characteristics of phosphorylated peptides: phosphorylated proteins show decreased hydrophobicity, ion suppression of phosphorylated peptides in the presence of non-phosphorylated peptides occurs and phosphorylated peptides are claimed to display reduced ionisation/detection efficiencies. Recently this was refuted by Steen *et al.* (145), who found no evidence for either of the aforementioned arguments using a set of synthetic phosphopeptides. They state that other, well-known, difficulties such as the dynamic nature of the phenomenon (equilibrium between kinase and

phosphatase activities) and substochiometric modification of peptides have a much larger negative influence on phosphopeptide detection. Finally, the fact that MS/MS fragmentation of serine- and threonine-phosphorylated peptide ions mainly results in loss of the labile phosphate group only can hamper their sequence analysis. This however, can also be used as will be discussed below.

Mass spectrometry-based phosphorylation analysis

A way to exploit the very specific loss of 98 Da (in the case of serine or threonine phosphorylation) is by performing a neutral loss scan experiment for which a triple quadrupole-type mass spectrometer can be used (146).

The phospho-specfic loss can also be applied for data dependent MS^3 switching. In such an experiment, the fragment that has lost phosphoric acid is automatically selected to undergo another round of CID. The peptide sequence is retrieved from the MS^3 spectrum and the site of phosphorylation can be determined using the specific mass of a dehydroalanine (in the case of phosphoserine) or dehydrobutyric acid (in the case of phosphothreonine) residue (147).

Hypothesis-driven mass spectrometry can also be used for the identification of phosphorylated peptides from a proteolytic digest (148,149). Theoretical masses of phosphopeptides are directly selected for CID enabling sequence and phosphorylation site determination. Even though this approach allows sequencing of low abundant phosphopeptides that would have been missed in a normal data-dependent MS run due to threshold settings, it can not be used for the phosphoproteomic analysis of complex mixtures, since in that case too many theoretical phosphorylation sites are present. Therefore, this approach is better suited for the analysis of purified proteins that contain specific kinase consensus sequences.

Analysis of phosphorylated peptides can also be performed using MALDI-TOF mass spectrometry. Optimisation of the detection of phosphopeptides can be achieved using additives like ammonium citrate (150,151). Differential enzymatic approaches for the detection of phosphopeptides have been applied for the analysis of purified proteins or for relatively simple protein mixtures (152). Upon phosphatase-catalyzed removal of phosphate from peptides, a mass shift of 80 Da towards lower m/z is observed, enabling the detection of phosphorylated peptides within a mixture.

Next to that, new mass spectrometric fragmentation methods are developed that induce fragmentation through electron capture (153,154) or electron transfer (155) methods, which mainly results in backbone

fragmentation. A big advantage of these approaches is that the phosphate group –or other possibly labile post-translation modifications- is not lost during analysis.

In general, chances of identifying phosphorylation sites on peptides from proteolytic digests are improved by reducing sample complexity. Even though antibodies against phosphorylated proteins have been successfully employed for tyrosine-phosphorylated proteins (156,157), the lack of antibodies specific for serine- and threonine-phosphorylated proteins has led to a very limited number of applications on these type of phosphorylated proteins (158). Therefore, other methods can be used for the analysis of S/T phosphorylation, such as (2D) gel- or liquid chromatography-based methods.

(2D) gel-based methods for protein phosphorylation analysis

A generally recognized advantage of 2D-GE is the ability to resolve protein isoforms and post-translationally modified proteins (159). In the case of protein phosphorylation the molecular weight of the protein hardly changes, while it reduces the iso-electric point (pI). This results in a horizontal shift of a protein spot towards lower pH upon phosphorylation of the protein, often detected as trains of spots in the case of multiple phosphorylation sites. The presence of such trains indicates that a protein might be phosphorylated. Other modifications however, like lysine or N-terminal acetylation have a similar effect on protein mass and pI. Therefore a number of methods that enable specific detection of phosphorylated proteins in gel have been developed. A conventional method for the direct qualitative and quantitative analysis of protein phosphorylation is through *in vitro* labelling using ^{32}P-γ-ATP. After labelling, proteins are resolved by 2D-GE and changes in phosphorylation are calculated using spot intensities in autoradiograms. The use of radioactively labelled phosphate however, is hazardous and contaminates analytical equipment and is therefore not ideally suited for phosphoproteomic analyses.

Recently, a fluorescent stain was introduced that specifically visualizes (in gel separated) phosphorylated proteins through metal coordination. The characteristics of the staining method enable multiplexed proteomics experiments in which both general protein expression as well as protein phosphorylation can be studied with general fluorescent protein stains (such as SYPRO, described above) and the phospho-specific fluorescent stain, respectively (160-163). An example of a gel on which phosphorylated proteins were visualised using this fluorescent staining method is given in Figure 8. Phosphorylated proteins of interest can be excised from the gel and identified using mass spectrometry. The spot indicated in Figure 8 was identified as nucleophosmin, a nuclear protein known to be phosphorylated on at least one serine and one

threonine residue (164). Since the phosphate remains on the protein during the analysis, proteolytic digestion and subsequent liquid chromatography-tandem mass spectrometry enabled the identification one of the phosphopeptides and the site of phosphorylation. In general however, fluorescence detection is more sensitive than mass spectrometry and as a result identification can be problematic, and phosphosite identification usually requires additional enrichment steps.

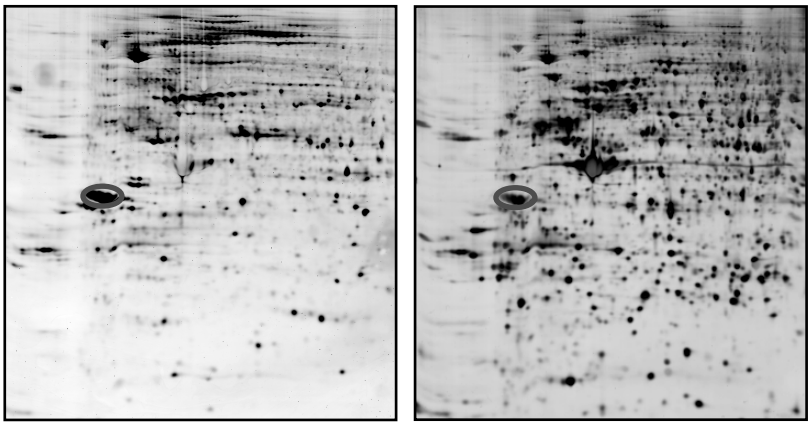


Figure 8. *Example of a 13 cm 3-10NL two dimensional gel of a nuclear lysate that was first stained with ProQ Diamond phosphorylation-specific fluorescent staining and subsequently with silver. The indicated protein on the acidic (left) side of the gel was identified as nucleophosmin, which is known to be phosphorylated at at least two sites.*

LC-based methods for protein phosphorylation analysis

In parallel to the progress in detection methods for phosphorylated proteins in gel, efforts have been made to develop column materials suitable for (multidimensional) liquid chromatography that specifically bind phosphorylated peptides and therefore can be used for enrichment of this type of peptides.

An example is immobilized metal affinity chromatography (IMAC), which makes use of metal coordination of negatively charged phosphate groups on peptides. A trivalent metal ion, like Fe^{III} or Ga^{III}, is immobilized to a negatively charged matrix (imidodiacetate (IDA) or nitrilotriacetate (NTA)) and is subsequently used to specifically enrich phosphorylated peptides through coordination binding (152,165-173). Bound phospho-peptides can be eluted using either basic or phosphate buffers. A disadvantage of IMAC is the fact that acidic

amino acid residues, like glutamic and aspartic acid, also carry a negatively charged side chains. This causes aspecific binding of acidic peptides, which in turn leads to a decrease in sensitivity for the phosphorylated peptides. Carboxymethylation of glutamic and aspartic acid side chains, to eliminate the negative charge, has been described to decrease the aspecfic binding of these peptides (171), but is not ideal due to the number of side reactions that can occur and to the fact that the reaction is reversible and therefore does not run to completion.

In 2004, Pinkse *et al* (174) introduced a new column material for the enrichment of phosphorylated peptides. This titanium oxide material was successfully used to enrich phosphorylated peptides from a digested autophosphorylated kinase. The mechanism of binding specificity is not exactly known, but it is assumed that coordination of the phosphorylated peptides by metal plays a crucial role. The large advantage over IMAC was reported to be the fact that no metal loading and associated additional washing steps are needed, thereby limiting analysis time.

Further improvement can be obtained by multidimensional chromatographic approaches in which the IMAC or titanium oxide step is preceded by ion exchange chromatography. In the case of strong anion exchange chromatography, this reduces sample complexity but does not solve the problem of aspecific binding of acidic peptides (175). Using strong cation exchange (SCX) chromatography at low pH however, allowed enrichment of phosphorylated peptides from a HeLa cell lysate (147). At low pH, most (tryptic) peptides carry two positive charges (on the N-terminal and lysine side chain amino group), while a phosphorylated peptide only carries one, due to the negative charge of the phosphate group. Consequently, phosphorylated peptides bind relatively weak to the SCX column, which can be used to elute phosphorylated peptides in the isocratic part of the gradient, together with N-terminally acetylated and C-terminal peptides. The other peptides will elute during the salt gradient. The first fractions eluting from the ion exchange column can be subsequently injected onto a titanium oxide column to further enrich phosphorylated peptides.

Chemical methods for protein phosphorylation analysis

Besides the physical properties of phosphorylated peptides, the well-known fact that serine- and threonine-phosphorylated peptides are susceptible to base-catalysed β-elimination is frequently used for the enrichment of phosphorylated peptides. The reaction, of which the mechanism is shown in Figure 9, is catalyzed by dilute alkali in the presence of group II metal ions.

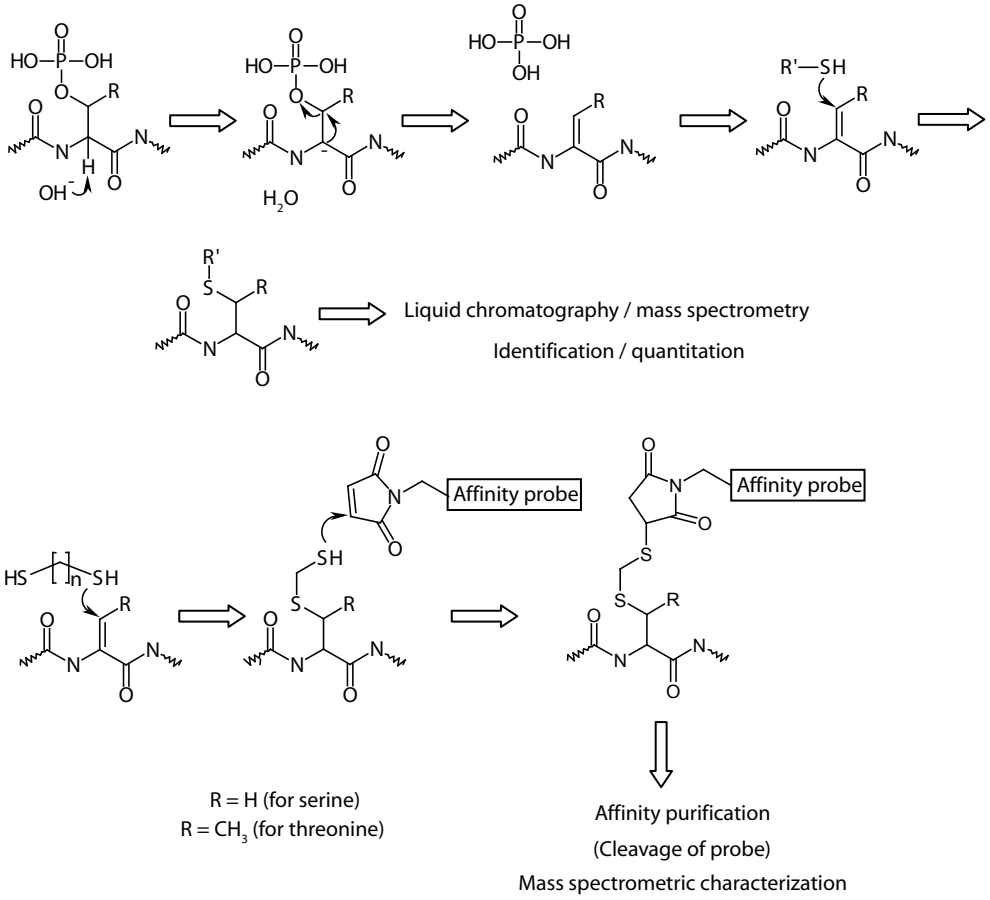


Figure 9. *General reaction mechanism for the β-elimination, Michael addition and subsequent affinity purification of phosphorylated serine- and threonine-containing peptides.*

The occurrence of an unwanted side reaction during the reaction, *i.e.* the elimination of glycosidic groups from serine or threonine residues is two orders of magnitude slower than that of phosphate groups using barium hydroxide (176). Therefore, barium hydroxide has been described to be most suited for this purpose. Elimination results in the formation of a dehydro amino acid (dehydroalanine in the case of phosphorylated serine and dehydrobutyric acid in the case of a phosphorylated threonine) of which the α, β-unsaturated side chain is susceptible to nucleophilic attack. This opens up a large set of reactions that can be carried out to modify the formerly phosphorylated peptides (177), see

Figure 9. First of all, different aliphatic thiols have been described to positively influence the chromatographic behaviour of phosphopeptides (178-180) and their stable isotope labelled forms allow relative quantitation of phosphorylation states (181-183). A specific example is the nucleophilic addition of cystamine (H_2N-C_2H_4-SH) to the dehydro amino acid which not only introduces an extra amino group resulting in improved ionisation of the formerly phosphorylated peptide (184), but with the newly formed structure being a lysine analogue this also introduces a 'phosphospecific' proteolytic (trypsin, LysC) cleavage site (185-187).

More frequently however, dithiols are used as nucleophiles: upon addition of a dithiol, an sulfhydryl group is introduced in the peptide that can be used to enable thiol-based affinity purification (188) or as a handle for further modification, for example with biotin to allow affinity purification of modified phosphopeptides (189). Increasingly advanced approaches, such as the PhIAT approach by Goshe (190-192) have been reported. This approach was further optimised with the introduction of labile linkers in a biotin probe that can be used to cleave off the biotin moiety directly after affinity purification and prior to mass spectrometric analysis (193,194). This is an advantage, since biotin is known to hamper sequence analysis due to its own fragmentation in MS/MS experiments. Instead of using affinity couples to modify the functionalised peptide, the sulfhydryl group can also be used for fluorescent labelling of the peptide (195,196).

Latest developments in chemical phosphoproteomic approaches involve modification of sulfhydryl-functionalized peptides with solid-phase resin coupled probes (197-200) and the use of dendrimers for the selective enrichment of phosphorylated peptides (201).

As mentioned before, a disadvantage of the chemistry described here is that not only phosphate, but also other functional groups, such as glycosidic esters are susceptible to β-elimination. This means that final reaction products can originate from phosphorylated or glycosylated peptides or even from peptides carrying other modifications on serine and threonine hydroxyl groups. GlcNAc modification of S and T residues occurs frequently (202-204) and several studies have described interplay between this type of modification and phosphorylation (205-208). Therefore, distinguishing between both is important. An approach can be the combination of chemistry and an enzymatic treatment in which GlcNAc groups are removed using glycanases. In this way both glycosylation and phosphorylation can be subsequently analysed.

V. *Chemical proteomics*

The genomic revolution has provided insight into the functioning of cells and the important role of proteins therein. However, understanding protein function within complex cellular networks is needed to enable, for example, the discovery of novel drug targets. Even though the present proteomics methods are able to resolve complex mixtures of proteins, they are often challenged by the dynamics of the proteome.

In an attempt to focus proteomic efforts on subsets of physiologically important protein targets, activity-based proteomics (or chemical proteomics) has been initiated. This approach makes use of small molecules called activity-based probes (ABPs) that are used to tag, enrich, and isolate, particular sets of proteins based on their enzymatic activity analogous to chemical phosphoproteomic approaches described above.

Chemical probes can be customized to react with different (enzymatic) targets through the use of chemically reactive 'warhead' groups, coupled to selective binding elements that control their overall specificity, thereby providing an insight into enzymatic activity within complex proteomes (209, 210). In this way, (additional) targets (and their interaction partners) of a particular ligand can be identified. Ligands can be other proteins, peptides or chemical compounds, like inhibitors or cofactors.

The general strategy of a chemical proteomics experiment (211) starts with the modification and immobilization of a ligand of interest onto solid phase. Examples of this are the immobilization of microcystin, a bacterial toxin, through nucleophilic addition of *L*-cystein to Microcystin-LR and the subsequent coupling of that compound to activated CH Sepharose 4B (212), or the coupling of amino-modified cyclin-dependent kinase inhibitors onto carbonyldiimidazole-activated agarose beads (213,214).

During these immobilization procedures it is very important to ensure that the interaction between the proteins of interest and the immobilized ligand is not lost as a result of chemical modification and/or coupling to solid phase. Varying linker molecules used for immobilization or modifying other functional groups within the compound of interest can circumvent possible problems at this level. Subsequently, the solid phase bound inhibitor can be used in a cellular lysate, to purify its (known and unknown) targets (215) and thereby gain insight into in the molecular mechanisms of known drugs. Depending on the experimental set-up co-purifications of proteins binding to the actual targets of the immobilized compound can be pulled down and analysed as well. The material used to immobilize a ligand on can also show affinity for proteins in the lysate and therefore multiple controls, including unmodified beads or beads onto

which an inactive analogue of the compound of interest was immobilized, should be included in the analysis. To get round problems with aspecific binding, several elution methods can be applied in which for example salt or detergents are used that wash away aspecifically binding proteins. On the other hand, competitive elution with the free ligand is the most specific elution method.

Examples of chemical proteomics approaches identified malate dehydrogenase (mitochondrial, but not cytoplasmic) as major paullone-binding protein, which is an inhibitor of cyclin-dependent kinases (215) or revealed alternative cellular modes of action for gefitinib, an epidermal growth factor receptor kinase inhibitor (216). From the set of interacting proteins found in experiment in which a complete protein, recombinant O^6-methylguanine-DNA methyltransferase, was immobilized it was concluded that this protein not only functions in DNA repair, but integrates this process with other cellular events, such as replication and cell cycle progression (217).

Analogous approaches can be designed in which peptides with a certain motif are immobilized to see which proteins bind to that peptide, like for instance the use of an immobilized polyproline peptide to enrich for SH3 domain-containing proteins, or the selective purification of DNA binding proteins, such as transcription factors, using immobilized DNA (218).

To fully utilize the power of chemical proteomic methods high throughput methods are required. Therefore libraries are built to perform large-scale screens with (219) that can aid the pharmaceutical industry to a great extent by determining protein targets for bioactive compounds.

VI. *Scope of this thesis*

This thesis describes how various approaches in proteomic research were developed and used to investigate the complex protein networks that underlie the cellular response to double strand DNA breaks of human lymphoblastoid cells. In quantitative proteomics accurate and reproducible quantitation methods are essential. Therefore two state-of-the-art techniques used for the quantitation of protein expression levels, i.e. two dimensional difference in-gel electrophoresis (2D-DiGE) and metabolic stable isotope labelling were compared, which is described in Chapter 2. Both methods were applied and compared in a single quantitative proteomic experiment in which differences in protein expression levels of the yeast *Saccharomyces cervisiae* grown under nitrogen- and carbon-limited conditions in chemostat cultures were determined. Subsequently, 2D-DiGE was applied to investigate the cellular response to bleomycin-induced DNA damage in human lymphoblastoid cell lines. The results of this experiment are

described in Chapter 3, which illustrates that a proteomics approach can be used for the accurate quantitation of protein expression levels.

The dataset obtained in Chapter 3 was further analysed to reveal differential protein expression between individuals that show different sensitivities towards bleomycin-induced DNA damage. The results, described in Chapter 4, demonstrate that these individuals show different expression patterns for certain proteins. This sheds new light on the biochemical mechanisms underlying the increased sensitivity towards DNA damage and can aid in the development of cancer therapies.

Numerous cellular processes are not only regulated by gene expression or protein degradation. Post-translational modification of certain amino acids in proteins can also influence their activity and thus that of the signalling pathways in which they function. One of the most important post-translational modifications is protein phosphorylation on serine, threonine and tyrosine residues. In order to focus on the influence of this modification a novel chemical proteomics approach for the enrichment and mass spectrometric analysis of serine- and threonine-phosphorylated peptides was developed, which is described in Chapter 5. This method makes use of β-elimination of the phosphate moiety and the subsequent (Michael) addition of a dithiol to which a probe consisting of a biotin moiety, acid cleavable linker and a maleimid-functionalized endgroup, is coupled.

Since DNA is densely packed in chromatin when stored in nuclei, DNA-related processes such as transcription but also DNA repair require relaxation of chromatin to increase accessibility of transcription and/or repair factors to particular sites on DNA. How these remodelling events are regulated is largely unknown. Therefore an LC-MS-based analysis of phosphorylation of histone-binding proteins before and after DNA damage induction was peformed, which is described in Chapter 6. Histone-binding proteins were enriched using an immobilized histone-mimicking peptide. Following in solution dual digestion, the peptide mixtures were analysed using multidimensional nano LC (TiO_2/SCX/RP) coupled to Fourier transform mass spectrometry. This analysis provides improved insight into the role of interactions between proteins and core histone N-terminal tails in the regulation of the response to DNA double strand breaks.

Taken together, we have combined quantitative proteomics of cellular lysates, subproteomics of proteins carrying phosphorylation (phosphoproteomics) and the analysis of the subgroup of proteins that bind to chromatin in an attempt to map the complex mechanisms that are triggered in human cell lines upon DNA damage induction.

REFERENCES

1. De Bont, R., and van Larebeke, N. (2004) Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis* **19**, 169-185
2. Sancar, A., Lindsey-Boltz, L. A., Unsal-Kacmaz, K., and Linn, S. (2004) Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu Rev Biochem* **73**, 39-85
3. Richardson, C., Horikoshi, N., and Pandita, T. K. (2004) The role of the DNA double-strand break response network in meiosis. *DNA Repair (Amst)* **3**, 1149-1164
4. Krangel, M. S. (2003) Gene segment selection in V(D)J recombination: accessibility and beyond. *Nat Immunol* **4**, 624-630
5. Rouse, J., and Jackson, S. P. (2002) Interfaces between the detection, signaling, and repair of DNA damage. *Science* **297**, 547-551
6. Zhou, B. B., and Elledge, S. J. (2000) The DNA damage response: putting checkpoints in perspective. *Nature* **408**, 433-439
7. McGowan, C. H., and Russell, P. (2004) The DNA damage response: sensing and signaling. *Curr Opin Cell Biol* **16**, 629-633
8. Li, L., and Zou, L. (2005) Sensing, signaling, and responding to DNA damage: organization of the checkpoint pathways in mammalian cells. *J Cell Biochem* **94**, 298-306
9. Shiloh, Y. (2001) ATM (ataxia telangiectasia mutated): expanding roles in the DNA damage response and cellular homeostasis. *Biochem Soc Trans* **29**, 661-666
10. Downs, J. A., and Jackson, S. P. (2004) A means to a DNA end: the many roles of Ku. *Nat Rev Mol Cell Biol* **5**, 367-378
11. Chu, G. (1996) Role of the Ku autoantigen in V(D)J recombination and double-strand break repair. *Curr Top Microbiol Immunol* **217**, 113-132
12. Wang, H., Guan, J., Perrault, A. R., Wang, Y., and Iliakis, G. (2001) Replication protein A2 phosphorylation after DNA damage by the coordinated action of ataxia telangiectasia-mutated and DNA-dependent protein kinase. *Cancer Res* **61**, 8554-8563
13. Balajee, A. S., and Geard, C. R. (2004) Replication protein A and gamma-H2AX foci assembly is triggered by cellular response to DNA double-strand breaks. *Exp Cell Res* **300**, 320-334
14. Jackson, S. P. (2002) Sensing and repairing DNA double-strand breaks. *Carcinogenesis* **23**, 687-696
15. Iliakis, G., Wang, Y., Guan, J., and Wang, H. (2003) DNA damage checkpoint control in cells exposed to ionizing radiation. *Oncogene* **22**, 5834-5847
16. Kastan, M. B., and Lim, D. S. (2000) The many substrates and functions of ATM. *Nat Rev Mol Cell Biol* **1**, 179-186
17. Durocher, D., and Jackson, S. P. (2001) DNA-PK, ATM and ATR as sensors of DNA damage: variations on a theme? *Curr Opin Cell Biol* **13**, 225-231
18. Abraham, R. T. (2001) Cell cycle checkpoint signaling through the ATM and ATR kinases. *Genes Dev* **15**, 2177-2196
19. Kastan, M. B., Lim, D. S., Kim, S. T., and Yang, D. (2001) ATM--a key determinant of multiple cellular responses to irradiation. *Acta Oncol* **40**, 686-688
20. Smith, G. C., and Jackson, S. P. (1999) The DNA-dependent protein kinase. *Genes Dev* **13**, 916-934
21. Smith, G. C., Divecha, N., Lakin, N. D., and Jackson, S. P. (1999) DNA-dependent protein kinase and related proteins. *Biochem Soc Symp* **64**, 91-104
22. Nojima, H. (2004) G1 and S-phase checkpoints, chromosome instability, and cancer. *Methods Mol Biol* **280**, 3-49
23. Bartek, J., and Lukas, J. (2001) Mammalian G1- and S-phase checkpoints in response to DNA damage. *Curr Opin Cell Biol* **13**, 738-747
24. Bartek, J., and Lukas, J. (2001) Pathways governing G1/S transition and their response to DNA damage. *FEBS Lett* **490**, 117-122
25. Cobb, J. A., Shimada, K., and Gasser, S. M. (2004) Redundancy, insult-specific sensors and thresholds: unlocking the S-phase checkpoint response. *Curr Opin Genet Dev* **14**, 292-300

26. Kanaar, R., Hoeijmakers, J. H., and van Gent, D. C. (1998) Molecular mechanisms of DNA double strand break repair. *Trends Cell Biol* **8**, 483-489
27. Norbury, C. J., and Zhivotovsky, B. (2004) DNA damage-induced apoptosis. *Oncogene* **23**, 2797-2808
28. Wyman, C., Ristic, D., and Kanaar, R. (2004) Homologous recombination-mediated double-strand break repair. *DNA Repair (Amst)* **3**, 827-833
29. Dudas, A., and Chovanec, M. (2004) DNA double-strand break repair by homologous recombination. *Mutat Res* **566**, 131-167
30. Powell, S. N., and Kachnic, L. A. (2003) Roles of BRCA1 and BRCA2 in homologous recombination, DNA replication fidelity and the cellular response to ionizing radiation. *Oncogene* **22**, 5784-5791
31. van den Bosch, M., Lohman, P. H., and Pastink, A. (2002) DNA double-strand break repair by homologous recombination. *Biol Chem* **383**, 873-892
32. Lees-Miller, S. P., and Meek, K. (2003) Repair of DNA double strand breaks by non-homologous end joining. *Biochimie* **85**, 1161-1173
33. Drouet, J., Delteil, C., Lefrancois, J., Concannon, P., Salles, B., and Calsou, P. (2005) DNA-dependent protein kinase and XRCC4-DNA ligase IV mobilization in the cell in response to DNA double strand breaks. *J Biol Chem* **280**, 7060-7069
34. Wolffe, A. P., and Guschin, D. (2000) Review: chromatin structural features and targets that regulate transcription. *J Struct Biol* **129**, 102-122
35. Celeste, A., Fernandez-Capetillo, O., Kruhlak, M. J., Pilch, D. R., Staudt, D. W., Lee, A., Bonner, R. F., Bonner, W. M., and Nussenzweig, A. (2003) Histone H2AX phosphorylation is dispensable for the initial recognition of DNA breaks. *Nat Cell Biol* **5**, 675-679
36. Foster, E. R., and Downs, J. A. (2005) Histone H2A phosphorylation in DNA double-strand break repair. *Febs J* **272**, 3231-3240
37. Stucki, M., Clapperton, J. A., Mohammad, D., Yaffe, M. B., Smerdon, S. J., and Jackson, S. P. (2005) MDC1 Directly Binds Phosphorylated Histone H2AX to Regulate Cellular Responses to DNA Double-Strand Breaks. *Cell* **123**, 1213-1226
38. Berardi, P., Russell, M., El-Osta, A., and Riabowol, K. (2004) Functional links between transcription, DNA repair and apoptosis. *Cell Mol Life Sci* **61**, 2173-2180
39. Huyen, Y., Zgheib, O., Ditullio Jr, R. A., Gorgoulis, V. G., Zacharatos, P., Petty, T. J., Sheston, E. A., Mellert, H. S., Stavridi, E. S., and Halazonetis, T. D. (2004) Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature* **432**, 406-411
40. Sanders, S. L., Portoso, M., Mata, J., Bahler, J., Allshire, R. C., and Kouzarides, T. (2004) Methylation of histone h4 lysine 20 controls recruitment of crb2 to sites of DNA damage. *Cell* **119**, 603-614
41. van Attikum, H., and Gasser, S. M. (2005) The histone code at DNA breaks: a guide to repair? *Nat Rev Mol Cell Biol*
42. Vidanes, G. M., Bonilla, C. Y., and Toczyski, D. P. (2005) Complicated tails: histone modifications and the DNA damage response. *Cell* **121**, 973-976
43. Hassa, P. O., and Hottiger, M. O. (2005) An epigenetic code for DNA damage repair pathways? *Biochem Cell Biol* **83**, 270-285
44. Liu, Y., and Kulesz-Martin, M. (2001) p53 protein at the hub of cellular DNA damage response pathways through sequence-specific and non-sequence-specific DNA binding. *Carcinogenesis* **22**, 851-860
45. Jimenez, G. S., Khan, S. H., Stommel, J. M., and Wahl, G. M. (1999) p53 regulation by post-translational modification and nuclear retention in response to diverse stresses. *Oncogene* **18**, 7656-7665
46. Bache, M., Pigorsch, S., Dunst, J., Wurl, P., Meye, A., Bartel, F., Schmidt, H., Rath, F. W., and Taubert, H. (2001) Loss of G2/M arrest correlates with radiosensitization in two human sarcoma cell lines with mutant p53. *Int J Cancer* **96**, 110-117
47. Oren, M. (1999) Regulation of the p53 tumor suppressor protein. *J Biol Chem* **274**, 36031-36034

48. Abraham, J., Kelly, J., Thibault, P., and Benchimol, S. (2000) Post-translational modification of p53 protein in response to ionizing radiation analyzed by mass spectrometry. *J Mol Biol* **295**, 853-864
49. Ashcroft, M., and Vousden, K. H. (1999) Regulation of p53 stability. *Oncogene* **18**, 7637-7643
50. Ashcroft, M., Kubbutat, M. H., and Vousden, K. H. (1999) Regulation of p53 function and stability by phosphorylation. *Mol Cell Biol* **19**, 1751-1758
51. Banin, S., Moyal, L., Shieh, S., Taya, Y., Anderson, C. W., Chessa, L., Smorodinsky, N. I., Prives, C., Reiss, Y., Shiloh, Y., and Ziv, Y. (1998) Enhanced phosphorylation of p53 by ATM in response to DNA damage. *Science* **281**, 1674-1677
52. Martinez, J. D., Craven, M. T., Joseloff, E., Milczarek, G., and Bowden, G. T. (1997) Regulation of DNA binding and transactivation in p53 by nuclear localization and phosphorylation. *Oncogene* **14**, 2511-2520
53. Gu, W., Luo, J., Brooks, C. L., Nikolaev, A. Y., and Li, M. (2004) Dynamics of the p53 acetylation pathway. *Novartis Found Symp* **259**, 197-205; discussion 205-197, 223-195
54. Luo, J., Li, M., Tang, Y., Laszkowska, M., Roeder, R. G., and Gu, W. (2004) Acetylation of p53 augments its site-specific DNA binding both in vitro and in vivo. *Proc Natl Acad Sci U S A* **101**, 2259-2264
55. Li, M., Luo, J., Brooks, C. L., and Gu, W. (2002) Acetylation of p53 inhibits its ubiquitination by Mdm2. *J Biol Chem* **277**, 50607-50611
56. Barlev, N. A., Liu, L., Chehab, N. H., Mansfield, K., Harris, K. G., Halazonetis, T. D., and Berger, S. L. (2001) Acetylation of p53 activates transcription through recruitment of coactivators/histone acetyltransferases. *Mol Cell* **8**, 1243-1254
57. Li, M., Chen, D., Shiloh, A., Luo, J., Nikolaev, A. Y., Qin, J., and Gu, W. (2002) Deubiquitination of p53 by HAUSP is an important pathway for p53 stabilization. *Nature* **416**, 648-653
58. Canman, C. E., Lim, D. S., Cimprich, K. A., Taya, Y., Tamai, K., Sakaguchi, K., Appella, E., Kastan, M. B., and Siliciano, J. D. (1998) Activation of the ATM kinase by ionizing radiation and phosphorylation of p53. *Science* **281**, 1677-1679
59. Morgan, S. E., and Kastan, M. B. (1997) p53 and ATM: cell cycle, cell death, and cancer. *Adv Cancer Res* **71**, 1-25
60. Cortez, D., Wang, Y., Qin, J., and Elledge, S. J. (1999) Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks. *Science* **286**, 1162-1166
61. Vorechovsky, I., Luo, L., Lindblom, A., Negrini, M., Webster, A. D., Croce, C. M., and Hammarstrom, L. (1996) ATM mutations in cancer families. *Cancer Res* **56**, 4130-4133
62. Sandoval, N., Platzer, M., Rosenthal, A., Dork, T., Bendix, R., Skawran, B., Stuhrmann, M., Wegner, R. D., Sperling, K., Banin, S., Shiloh, Y., Baumer, A., Bernthaler, U., Sennefelder, H., Brohm, M., Weber, B. H., and Schindler, D. (1999) Characterization of ATM gene mutations in 66 ataxia telangiectasia families. *Hum Mol Genet* **8**, 69-79
63. McKinnon, P. J. (2004) ATM and ataxia telangiectasia. *EMBO Rep* **5**, 772-776
64. Lavin, M. F., Khanna, K. K., Beamish, H., Teale, B., Hobson, K., and Watters, D. (1994) Defect in radiation signal transduction in ataxia-telangiectasia. *Int J Radiat Biol* **66**, S151-156
65. Barzilai, A., Rotman, G., and Shiloh, Y. (2002) ATM deficiency and oxidative stress: a new dimension of defective response to DNA damage. *DNA Repair (Amst)* **1**, 3-25
66. Ball, L. G., and Xiao, W. (2005) Molecular basis of ataxia telangiectasia and related diseases. *Acta Pharmacol Sin* **26**, 897-907
67. Lisby, M., Barlow, J. H., Burgess, R. C., and Rothstein, R. (2004) Choreography of the DNA damage response: spatiotemporal relationships among checkpoint and repair proteins. *Cell* **118**, 699-713
68. Khanna, K. K., and Jackson, S. P. (2001) DNA double-strand breaks: signaling, repair and the cancer connection. *Nat Genet* **27**, 247-254

69. Cloos, J., Spitz, M. R., Schantz, S. P., Hsu, T. C., Zhang, Z. F., Tobi, H., Braakhuis, B. J., and Snow, G. B. (1996) Genetic susceptibility to head and neck squamous cell carcinoma. *J Natl Cancer Inst* **88**, 530-535
70. Hsu, T. C. (1983) Genetic instability in the human population: a working hypothesis. *Hereditas* **98**, 1-9
71. Cloos, J., Nieuwenhuis, E. J., Boomsma, D. I., Kuik, D. J., van der Sterre, M. L., Arwert, F., Snow, G. B., and Braakhuis, B. J. (1999) Inherited susceptibility to bleomycin-induced chromatid breaks in cultured peripheral blood lymphocytes. *J Natl Cancer Inst* **91**, 1125-1130
72. Sieber, O. M., Heinimann, K., and Tomlinson, I. P. (2003) Genomic instability--the engine of tumorigenesis? *Nat Rev Cancer* **3**, 701-708
73. Cloos, J., Braakhuis, B. J., Steen, I., Copper, M. P., de Vries, N., Nauta, J. J., and Snow, G. B. (1994) Increased mutagen sensitivity in head-and-neck squamous-cell carcinoma patients, particularly those with multiple primary tumors. *Int J Cancer* **56**, 816-819
74. Cloos, J., Leemans, C. R., van der Sterre, M. L., Kuik, D. J., Snow, G. B., and Braakhuis, B. J. (2000) Mutagen sensitivity as a biomarker for second primary tumors after head and neck squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev* **9**, 713-717
75. Amundson, S. A., Xia, F., Wolfson, K., and Liber, H. L. (1993) Different cytotoxic and mutagenic responses induced by X-rays in two human lymphoblastoid cell lines derived from a single donor. *Mutat Res* **286**, 233-241
76. Bishay, K., Ory, K., Lebeau, J., Levalois, C., Olivier, M. F., and Chevillard, S. (2000) DNA damage-related gene expression as biomarkers to assess cellular response after gamma irradiation of a human lymphoblastoid cell line. *Oncogene* **19**, 916-923
77. Jen, K. Y., and Cheung, V. G. (2003) Transcriptional response of lymphoblastoid cells to ionizing radiation. *Genome Res* **13**, 2092-2100
78. Hsu, T. C., Shillitoe, E. J., Cherry, L. M., Lin, Q., Schantz, S. P., and Furlong, C. (1990) Cytogenetic characterization of 20 lymphoblastoid lines derived from human individuals differing in bleomycin sensitivity. *In Vitro Cell Dev Biol* **26**, 80-84
79. Adema, A. D., Cloos, J., Verheijen, R. H., Braakhuis, B. J., and Bryant, P. E. (2003) Comparison of bleomycin and radiation in the G2 assay of chromatid breaks. *Int J Radiat Biol* **79**, 655-661
80. Hecht, S. M. (2000) Bleomycin: new perspectives on the mechanism of action. *J Nat Prod* **63**, 158-168
81. Cloos, J., Gille, J. J., Steen, I., Lafleur, M. V., Retel, J., Snow, G. B., and Braakhuis, B. J. (1996) Influence of the antioxidant N-acetylcysteine and its metabolites on damage induced by bleomycin in PM2 bacteriophage DNA. *Carcinogenesis* **17**, 327-331
82. Conway, T., and Schoolnik, G. K. (2003) Microarray expression profiling: capturing a genome-wide portrait of the transcriptome. *Mol Microbiol* **47**, 879-889
83. Amundson, S. A., Bittner, M., Chen, Y., Trent, J., Meltzer, P., and Fornace, A. J., Jr. (1999) Fluorescent cDNA microarray hybridization reveals complexity and heterogeneity of cellular genotoxic stress responses. *Oncogene* **18**, 3666-3672
84. Amundson, S. A., and Fornace, A. J., Jr. (2001) Gene expression profiles for monitoring radiation exposure. *Radiat Prot Dosimetry* **97**, 11-16
85. Amundson, S. A., Bittner, M., Meltzer, P., Trent, J., and Fornace, A. J., Jr. (2001) Induction of gene expression as a monitor of exposure to ionizing radiation. *Radiat Res* **156**, 657-661
86. Amundson, S. A., and Fornace, A. J., Jr. (2003) Monitoring human radiation exposure by gene expression profiling: possibilities and pitfalls. *Health Phys* **85**, 36-42
87. Amundson, S. A., Grace, M. B., McLeland, C. B., Epperly, M. W., Yeager, A., Zhan, Q., Greenberger, J. S., and Fornace, A. J., Jr. (2004) Human in vivo radiation-induced biomarkers: gene expression changes in radiotherapy patients. *Cancer Res* **64**, 6368-6371
88. Tian, Q., Stepaniants, S. B., Mao, M., Weng, L., Feetham, M. C., Doyle, M. J., Yi, E. C., Dai, H., Thorsson, V., Eng, J., Goodlett, D., Berger, J. P., Gunter, B., Linseley, P. S., Stoughton, R. B., Aebersold, R., Collins, S. J., Hanlon, W. A., and Hood, L. E. (2004)

Integrated genomic and proteomic analyses of gene expression in Mammalian cells. *Mol Cell Proteomics* **3**, 960-969
89. Hack, C. J. (2004) Integrated transcriptome and proteome data: the challenges ahead. *Brief Funct Genomic Proteomic* **3**, 212-219
90. Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J. C., Yan, J. X., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L., and Hochstrasser, D. F. (1996) From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)* **14**, 61-65
91. O'Farrell, P. H. (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250**, 4007-4021
92. Rabilloud, T. (2000) Detecting proteins separated by 2-D gel electrophoresis. *Anal Chem* **72**, 48A-55A
93. Patton, W. F. (2002) Detection technologies in proteome analysis. *J Chromatogr B Analyt Technol Biomed Life Sci* **771**, 3-31
94. Szkanderova, S., Vavrova, J., Rezacova, M., Vokurkova, D., Pavlova, S., Smardova, J., and Stulik, J. (2003) Gamma irradiation results in phosphorylation of p53 at serine-392 in human T-lymphocyte leukaemia cell line MOLT-4. *Folia Biol (Praha)* **49**, 191-196
95. Szkanderova, S., Port, M., Stulik, J., Hernychova, L., Kasalova, I., Van Beuningen, D., and Abend, M. (2003) Comparison of the abundance of 10 radiation-induced proteins with their differential gene expression in L929 cells. *Int J Radiat Biol* **79**, 623-633
96. Szkanderova, S., Hernychova, L., Kasalova, I., Vavrova, J., Stulik, J., Abend, M., and van Beuningen, D. (2003) Proteomic analysis of radiation-induced alterations in L929 cells. *Folia Biol (Praha)* **49**, 15-25
97. Szkanderova, S., Vavrova, J., Hernychova, L., Neubauerova, V., Lenco, J., and Stulik, J. (2005) Proteome alterations in gamma-irradiated human T-lymphocyte leukemia cells. *Radiat Res* **163**, 307-315
98. Decker, E. D., Zhang, Y., Cocklin, R. R., Witzmann, F. A., and Wang, M. (2003) Proteomic analysis of differential protein expression induced by ultraviolet light radiation in HeLa cells. *Proteomics* **3**, 2019-2027
99. Karp, N. A., and Lilley, K. S. (2005) Maximising sensitivity for detecting changes in protein expression: Experimental design using minimal CyDyes. *Proteomics* **5**, 3105-3115
100. Lilley, K. S., Razzaq, A., and Dupree, P. (2002) Two-dimensional gel electrophoresis: recent advances in sample preparation, detection and quantitation. *Curr Opin Chem Biol* **6**, 46-50
101. Fey, S. J., and Larsen, P. M. (2001) 2D or not 2D. Two-dimensional gel electrophoresis. *Curr Opin Chem Biol* **5**, 26-33
102. Gorg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R., and Weiss, W. (2000) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **21**, 1037-1053
103. Rabilloud, T. (2002) Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. *Proteomics* **2**, 3-10
104. Yan, J. X., Devenish, A. T., Wait, R., Stone, T., Lewis, S., and Fowler, S. (2002) Fluorescence two-dimensional difference gel electrophoresis and mass spectrometry based proteomic analysis of Escherichia coli. *Proteomics* **2**, 1682-1698
105. 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
106. Unlu, M. (1999) Difference gel electrophoresis. *Biochem Soc Trans* **27**, 547-549
107. Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I., and Davison, M. (2001) Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* **1**, 377-396
108. Alban, A., David, S. O., Bjorkesten, L., Andersson, C., Sloge, E., Lewis, S., and Currie, I. (2003) A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics* **3**, 36-44

109. Dowsey, A. W., Dunn, M. J., and Yang, G. Z. (2003) The role of bioinformatics in two-dimensional gel electrophoresis. *Proteomics* **3**, 1567-1596
110. Kreil, D. P., Karp, N. A., and Lilley, K. S. (2004) DNA microarray normalization methods can remove bias from differential protein expression analysis of 2D difference gel electrophoresis results. *Bioinformatics* **20**, 2026-2034
111. Karp, N. A., Kreil, D. P., and Lilley, K. S. (2004) Determining a significant change in protein expression with DeCyder during a pair-wise comparison using two-dimensional difference gel electrophoresis. *Proteomics* **4**, 1421-1432
112. Karp, N. A., Griffin, J. L., and Lilley, K. S. (2005) Application of partial least squares discriminant analysis to two-dimensional difference gel studies in expression proteomics. *Proteomics* **5**, 81-90
113. Julka, S., and Regnier, F. (2004) Quantification in proteomics through stable isotope coding: a review. *J Proteome Res* **3**, 350-363
114. Goshe, M. B., and Smith, R. D. (2003) Stable isotope-coded proteomic mass spectrometry. *Curr Opin Biotechnol* **14**, 101-109
115. Ong, S. E., Foster, L. J., and Mann, M. (2003) Mass spectrometric-based approaches in quantitative proteomics. *Methods* **29**, 124-130
116. Gronborg, M., Kristiansen, T. Z., Iwahori, A., Chang, R., Reddy, R., Sato, N., Molina, H., Jensen, O. N., Hruban, R. H., Goggins, M. G., Maitra, A., and Pandey, A. (2005) Biomarker discovery from pancreatic cancer secretome using a differential proteomics approach. *Mol Cell Proteomics*
117. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**, 994-999
118. Stewart, II, Thomson, T., and Figeys, D. (2001) 18O labeling: a tool for proteomics. *Rapid Commun Mass Spectrom* **15**, 2456-2465
119. Smolka, M. B., Albuquerque, C. P., Chen, S. H., Schmidt, K. H., Wei, X. X., Kolodner, R. D., and Zhou, H. (2005) Dynamic Changes in Protein-Protein Interaction and Protein Phosphorylation Probed with Amine-reactive Isotope Tag. *Mol Cell Proteomics* **4**, 1358-1369
120. Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996) Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**, 466-469
121. Shevchenko, A., Wilm, M., Vorm, O., Jensen, O. N., Podtelejnikov, A. V., Neubauer, G., Mortensen, P., and Mann, M. (1996) A strategy for identifying gel-separated proteins in sequence databases by MS alone. *Biochem Soc Trans* **24**, 893-896
122. Mann, M., Hendrickson, R. C., and Pandey, A. (2001) Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem* **70**, 437-473
123. Yates, J. R., 3rd (2004) Mass spectral analysis in proteomics. *Annu Rev Biophys Biomol Struct* **33**, 297-316
124. Steen, H., and Mann, M. (2004) The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* **5**, 699-711
125. Hillenkamp, F., Karas, M., Beavis, R. C., and Chait, B. T. (1991) Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal Chem* **63**, 1193A-1203A
126. Guilhaus, M., Selby, D., and Mlynski, V. (2000) Orthogonal acceleration time-of-flight mass spectrometry. *Mass Spectrom Rev* **19**, 65-107
127. Whittal, R. M., and Li, L. (1995) High-resolution matrix-assisted laser desorption/ionization in a linear time-of-flight mass spectrometer. *Anal Chem* **67**, 1950-1954
128. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71
129. Kebarle, P. (2000) A brief overview of the present status of the mechanisms involved in electrospray mass spectrometry. *J Mass Spectrom* **35**, 804-817
130. Cech, N. B., and Enke, C. G. (2001) Practical implications of some recent studies in electrospray ionization fundamentals. *Mass Spectrom Rev* **20**, 362-387

131. Wilm, M., and Mann, M. (1996) Analytical properties of the nanoelectrospray ion source. *Anal Chem* **68**, 1-8
132. Whitehouse, C. M., Dreyer, R. N., Yamashita, M., and Fenn, J. B. (1985) Electrospray interface for liquid chromatographs and mass spectrometers. *Anal Chem* **57**, 675-679
133. Morris, H. R., Paxton, T., Dell, A., Langhorne, J., Berg, M., Bordoli, R. S., Hoyes, J., and Bateman, R. H. (1996) High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Commun Mass Spectrom* **10**, 889-896
134. Douglas, D. J., Frank, A. J., and Mao, D. (2005) Linear ion traps in mass spectrometry. *Mass Spectrom Rev* **24**, 1-29
135. Jonscher, K. R., and Yates, J. R., 3rd (1997) The quadrupole ion trap mass spectrometer--a small solution to a big challenge. *Anal Biochem* **244**, 1-15
136. Karas, M., and Hillenkamp, F. (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **60**, 2299-2301
137. Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., and Watanabe, C. (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci U S A* **90**, 5011-5015
138. James, P., Quadroni, M., Carafoli, E., and Gonnet, G. (1993) Protein identification by mass profile fingerprinting. *Biochem Biophys Res Commun* **195**, 58-64
139. Mann, M., Hojrup, P., and Roepstorff, P. (1993) Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol Mass Spectrom* **22**, 338-345
140. Pappin, D. J., Hojrup, P., and Bleasby, A. J. (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol* **3**, 327-332
141. Roepstorff, P., and Fohlman, J. (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* **11**, 601
142. Edelman, A. M., Blumenthal, D. K., and Krebs, E. G. (1987) Protein serine/threonine kinases. *Annu Rev Biochem* **56**, 567-613
143. Mann, M., Ong, S. E., Gronborg, M., Steen, H., Jensen, O. N., and Pandey, A. (2002) Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol* **20**, 261-268
144. Cans, C., Mangano, R., Barila, D., Neubauer, G., and Superti-Furga, G. (2000) Nuclear tyrosine phosphorylation: the beginning of a map. *Biochem Pharmacol* **60**, 1203-1215
145. Steen, H., Jebanathirajah, J. A., Rush, J., Morrice, N., and Kirschner, M. W. (2005) Phosphorylation analysis by mass spectrometry: Myths, facts and the consequences for qualitative and quantitative measurements. *Mol Cell Proteomics*
146. McLachlin, D. T., and Chait, B. T. (2001) Analysis of phosphorylated proteins and peptides by mass spectrometry. *Curr Opin Chem Biol* **5**, 591-602
147. Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villen, J., Li, J., Cohn, M. A., Cantley, L. C., and Gygi, S. P. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A* **101**, 12130-12135
148. Yi, Z., Luo, M., Carroll, C. A., Weintraub, S. T., and Mandarino, L. J. (2005) Identification of phosphorylation sites in insulin receptor substrate-1 by hypothesis-driven high-performance liquid chromatography-electrospray ionization tandem mass spectrometry. *Anal Chem* **77**, 5693-5699
149. Chang, E. J., Archambault, V., McLachlin, D. T., Krutchinsky, A. N., and Chait, B. T. (2004) Analysis of protein phosphorylation by hypothesis-driven multiple-stage mass spectrometry. *Anal Chem* **76**, 4472-4483
150. Asara, J. M., and Allison, J. (1999) Enhanced detection of phosphopeptides in matrix-assisted laser desorption/ionization mass spectrometry using ammonium salts. *J Am Soc Mass Spectrom* **10**, 35-44
151. Yang, X., Wu, H., Kobayashi, T., Solaro, R. J., and van Breemen, R. B. (2004) Enhanced ionization of phosphorylated peptides during MALDI TOF mass spectrometry. *Anal Chem* **76**, 1532-1536

152. Stensballe, A., Andersen, S., and Jensen, O. N. (2001) Characterization of phosphoproteins from electrophoretic gels by nanoscale Fe(III) affinity chromatography with off-line mass spectrometry analysis. *Proteomics* **1**, 207-222
153. Shi, S. D., Hemling, M. E., Carr, S. A., Horn, D. M., Lindh, I., and McLafferty, F. W. (2001) Phosphopeptide/phosphoprotein mapping by electron capture dissociation mass spectrometry. *Anal Chem* **73**, 19-22
154. Stensballe, A., Jensen, O. N., Olsen, J. V., Haselmann, K. F., and Zubarev, R. A. (2000) Electron capture dissociation of singly and multiply phosphorylated peptides. *Rapid Commun Mass Spectrom* **14**, 1793-1800
155. Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* **101**, 9528-9533
156. Rush, J., Moritz, A., Lee, K. A., Guo, A., Goss, V. L., Spek, E. J., Zhang, H., Zha, X. M., Polakiewicz, R. D., and Comb, M. J. (2005) Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat Biotechnol* **23**, 94-101
157. Blagoev, B., Ong, S. E., Kratchmarova, I., and Mann, M. (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat Biotechnol* **22**, 1139-1145
158. Gronborg, M., Kristiansen, T. Z., Stensballe, A., Andersen, J. S., Ohara, O., Mann, M., Jensen, O. N., and Pandey, A. (2002) A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate. *Mol Cell Proteomics* **1**, 517-527
159. Gorg, A., Weiss, W., and Dunn, M. J. (2004) Current two-dimensional electrophoresis technology for proteomics. *Proteomics* **4**, 3665-3685
160. Martin, K., Steinberg, T. H., Cooley, L. A., Gee, K. R., Beechem, J. M., and Patton, W. F. (2003) Quantitative analysis of protein phosphorylation status and protein kinase activity on microarrays using a novel fluorescent phosphorylation sensor dye. *Proteomics* **3**, 1244-1255
161. Steinberg, T. H., Agnew, B. J., Gee, K. R., Leung, W. Y., Goodman, T., Schulenberg, B., Hendrickson, J., Beechem, J. M., Haugland, R. P., and Patton, W. F. (2003) Global quantitative phosphoprotein analysis using Multiplexed Proteomics technology. *Proteomics* **3**, 1128-1144
162. Schulenberg, B., Goodman, T. N., Aggeler, R., Capaldi, R. A., and Patton, W. F. (2004) Characterization of dynamic and steady-state protein phosphorylation using a fluorescent phosphoprotein gel stain and mass spectrometry. *Electrophoresis* **25**, 2526-2532
163. Unwin, R. D., Sternberg, D. W., Lu, Y., Pierce, A., Gilliland, D. G., and Whetton, A. D. (2005) Global effects of BCR/ABL and TEL/PDGFRbeta expression on the proteome and phosphoproteome: identification of the Rho pathway as a target of BCR/ABL. *J Biol Chem* **280**, 6316-6326
164. Maiguel, D. A., Jones, L., Chakravarty, D., Yang, C., and Carrier, F. (2004) Nucleophosmin sets a threshold for p53 response to UV radiation. *Mol Cell Biol* **24**, 3703-3711
165. Gaberc-Porekar, V., and Menart, V. (2001) Perspectives of immobilized-metal affinity chromatography. *J Biochem Biophys Methods* **49**, 335-360
166. Posewitz, M. C., and Tempst, P. (1999) Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal Chem* **71**, 2883-2892
167. Cao, P., and Stults, J. T. (2000) Mapping the phosphorylation sites of proteins using on-line immobilized metal affinity chromatography/capillary electrophoresis/electrospray ionization multiple stage tandem mass spectrometry. *Rapid Commun Mass Spectrom* **14**, 1600-1606
168. Cao, P., and Stults, J. T. (1999) Phosphopeptide analysis by on-line immobilized metal-ion affinity chromatography-capillary electrophoresis-electrospray ionization mass spectrometry. *J Chromatogr A* **853**, 225-235
169. Ahn, Y. H., Park, E. J., Cho, K., Kim, J. Y., Ha, S. H., Ryu, S. H., and Yoo, J. S. (2004) Dynamic identification of phosphopeptides using immobilized metal ion affinity

chromatography enrichment, subsequent partial beta-elimination/chemical tagging and matrix-assisted laser desorption/ionization mass spectrometric analysis. *Rapid Commun Mass Spectrom* **18**, 2495-2501

170. Ficarro, S., Chertihin, O., Westbrook, V. A., White, F., Jayes, F., Kalab, P., Marto, J. A., Shabanowitz, J., Herr, J. C., Hunt, D., and Visconti, P. E. (2002) Phosphoproteome analysis of human sperm. Evidence of tyrosine phosphorylation of AKAP 3 and valosin containing protein/P97 during capacitation. *J Biol Chem*
171. Ficarro, S. B., McCleland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F., and White, F. M. (2002) Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat Biotechnol* **20**, 301-305
172. Riggs, L., Sioma, C., and Regnier, F. E. (2001) Automated signature peptide approach for proteomics. *J Chromatogr A* **924**, 359-368
173. Hart, S. R., Waterfield, M. D., Burlingame, A. L., and Cramer, R. (2002) Factors governing the solubilization of phosphopeptides retained on ferric NTA IMAC beads and their analysis by MALDI TOFMS. *J Am Soc Mass Spectrom* **13**, 1042-1051
174. Pinkse, M. W., Uitto, P. M., Hilhorst, M. J., Ooms, B., and Heck, A. J. (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem* **76**, 3935-3943
175. Nuhse, T. S., Stensballe, A., Jensen, O. N., and Peck, S. C. (2003) Large-scale Analysis of in Vivo Phosphorylated Membrane Proteins by Immobilized Metal Ion Affinity Chromatography and Mass Spectrometry. *Mol Cell Proteomics* **2**, 1234-1243
176. Byford, M. F. (1991) Rapid and selective modification of phosphoserine residues catalysed by Ba2+ ions for their detection during peptide microsequencing. *Biochem J* **280 (Pt 1)**, 261-265
177. Salih, E. (2005) Phosphoproteomics by mass spectrometry and classical protein chemistry approaches. *Mass Spectrom Rev* **24**, 828-846
178. Molloy, M. P., and Andrews, P. C. (2001) Phosphopeptide derivatization signatures to identify serine and threonine phosphorylated peptides by mass spectrometry. *Anal Chem* **73**, 5387-5394
179. Mega, T., Hamazume, Y., Hong, Y. M., and Ikenaka, T. (1986) Studies on the methods for the determination of phosphorylation sites in highly phosphorylated peptides or proteins: phosphorylation sites of hen egg white riboflavin binding protein. *J Biochem (Tokyo)* **100**, 1109-1116
180. Holmes, C. F. (1987) A new method for the selective isolation of phosphoserine-containing peptides. *FEBS Lett* **215**, 21-24
181. Glinski, M., Romeis, T., Witte, C. P., Wienkoop, S., and Weckwerth, W. (2003) Stable isotope labeling of phosphopeptides for multiparallel kinase target analysis and identification of phosphorylation sites. *Rapid Commun Mass Spectrom* **17**, 1579-1584
182. Jaffe, H., Veeranna, and Pant, H. C. (1998) Characterization of serine and threonine phosphorylation sites in beta-elimination/ethanethiol addition-modified proteins by electrospray tandem mass spectrometry and database searching. *Biochemistry* **37**, 16211-16224
183. Weckwerth, W., Willmitzer, L., and Fiehn, O. (2000) Comparative quantification and identification of phosphoproteins using stable isotope labeling and liquid chromatography/mass spectrometry. *Rapid Commun Mass Spectrom* **14**, 1677-1681
184. Steen, H., and Mann, M. (2002) A new derivatization strategy for the analysis of phosphopeptides by precursor ion scanning in positive ion mode. *J Am Soc Mass Spectrom* **13**, 996-1003
185. Knight, Z. A., Schilling, B., Row, R. H., Kenski, D. M., Gibson, B. W., and Shokat, K. M. (2003) Phosphospecific proteolysis for mapping sites of protein phosphorylation. *Nat Biotechnol* **21**, 1047-1054
186. Peng, J. (2004) A chemically designed enzymatic cleavage site for phosphoproteome analysis. *Chembiochem* **5**, 768-770

187. McCormick, D. J., Holmes, M. W., Muddiman, D. C., and Madden, B. J. (2005) Mapping sites of protein phosphorylation by mass spectrometry utilizing a chemical-enzymatic approach: characterization of products from alpha-S1 casein phosphopeptides. *J Proteome Res* **4**, 424-434
188. Amoresano, A., Marino, G., Cirulli, C., and Quemeneur, E. (2004) Mapping phosphorylation sites: a new strategy based on the use of isotopically labelled DTT and mass spectrometry. *Eur J Mass Spectrom (Chichester, Eng)* **10**, 401-412
189. Oda, Y., Nagasu, T., and Chait, B. T. (2001) Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat Biotechnol* **19**, 379-382
190. 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. *Anal Chem* **74**, 607-616
191. Goshe, M. B., Conrads, T. P., Panisko, E. A., Angell, N. H., Veenstra, T. D., and Smith, R. D. (2001) Phosphoprotein isotope-coded affinity tag approach for isolating and quantitating phosphopeptides in proteome-wide analyses. *Anal Chem* **73**, 2578-2586
192. Qian, W. J., Goshe, M. B., Camp, D. G., 2nd, Yu, L. R., Tang, K., and Smith, R. D. (2003) Phosphoprotein isotope-coded solid-phase tag approach for enrichment and quantitative analysis of phosphopeptides from complex mixtures. *Anal Chem* **75**, 5441-5450
193. Adamczyk, M., Gebler, J. C., and Wu, J. (2001) Selective analysis of phosphopeptides within a protein mixture by chemical modification, reversible biotinylation and mass spectrometry. *Rapid Commun Mass Spectrom* **15**, 1481-1488
194. van der Veken, P., Dirksen, E. H., Ruijter, E., Elgersma, R. C., Heck, A. J., Rijkers, D. T., Slijper, M., and Liskamp, R. M. (2005) Development of a Novel Chemical Probe for the Selective Enrichment of Phosphorylated Serine- and Threonine-Containing Peptides. *Chembiochem* **6,** 2271-2280
195. Fadden, P., and Haystead, T. A. (1995) Quantitative and selective fluorophore labeling of phosphoserine on peptides and proteins: characterization at the attomole level by capillary electrophoresis and laser-induced fluorescence. *Anal Biochem* **225**, 81-88
196. Stevens, S. M., Jr., Chung, A. Y., Chow, M. C., McClung, S. H., Strachan, C. N., Harmon, A. C., Denslow, N. D., and Prokai, L. (2005) Enhancement of phosphoprotein analysis using a fluorescent affinity tag and mass spectrometry. *Rapid Commun Mass Spectrom* **19**, 2157-2162
197. Chowdhury, S. M., Munske, G. R., Siems, W. F., and Bruce, J. E. (2005) A new maleimide-bound acid-cleavable solid-support reagent for profiling phosphorylation. *Rapid Commun Mass Spectrom* **19**, 899-909
198. Tseng, H. C., Ovaa, H., Wei, N. J., Ploegh, H., and Tsai, L. H. (2005) Phosphoproteomic analysis with a solid-phase capture-release-tag approach. *Chem Biol* **12**, 769-777
199. McLachlin, D. T., and Chait, B. T. (2003) Improved beta-elimination-based affinity purification strategy for enrichment of phosphopeptides. *Anal Chem* **75**, 6826-6836
200. Thaler, F., Valsasina, B., Baldi, R., Xie, J., Stewart, A., Isacchi, A., Kalisz, H. M., and Rusconi, L. (2003) A new approach to phosphoserine and phosphothreonine analysis in peptides and proteins: chemical modification, enrichment via solid-phase reversible binding, and analysis by mass spectrometry. *Anal Bioanal Chem* **376**, 366-373
201. Tao, W. A., Wollscheid, B., O'Brien, R., Eng, J. K., Li, X. J., Bodenmiller, B., Watts, J. D., Hood, L., and Aebersold, R. (2005) Quantitative phosphoproteome analysis using a dendrimer conjugation chemistry and tandem mass spectrometry. *Nat Methods* **2**, 591-598
202. Zachara, N. E., and Hart, G. W. (2004) O-GlcNAc a sensor of cellular state: the role of nucleocytoplasmic glycosylation in modulating cellular function in response to nutrition and stress. *Biochim Biophys Acta* **1673**, 13-28
203. Chou, T. Y., and Hart, G. W. (2001) O-linked N-acetylglucosamine and cancer: messages from the glycosylation of c-Myc. *Adv Exp Med Biol* **491**, 413-418
204. Wells, L., Whelan, S. A., and Hart, G. W. (2003) O-GlcNAc: a regulatory post-translational modification. *Biochem Biophys Res Commun* **302**, 435-441

205. Whelan, S. A., and Hart, G. W. (2003) Proteomic approaches to analyze the dynamic relationships between nucleocytoplasmic protein glycosylation and phosphorylation. *Circ Res* **93**, 1047-1058
206. Liu, F., Iqbal, K., Grundke-Iqbal, I., Hart, G. W., and Gong, C. X. (2004) O-GlcNAcylation regulates phosphorylation of tau: a mechanism involved in Alzheimer's disease. *Proc Natl Acad Sci U S A* **101**, 10804-10809
207. Kamemura, K., and Hart, G. W. (2003) Dynamic interplay between O-glycosylation and O-phosphorylation of nucleocytoplasmic proteins: a new paradigm for metabolic control of signal transduction and transcription. *Prog Nucleic Acid Res Mol Biol* **73**, 107-136
208. Slawson, C., and Hart, G. W. (2003) Dynamic interplay between O-GlcNAc and O-phosphate: the sweet side of protein regulation. *Curr Opin Struct Biol* **13**, 631-636
209. Berger, A. B., Vitorino, P. M., and Bogyo, M. (2004) Activity-based protein profiling: applications to biomarker discovery, in vivo imaging and drug discovery. *Am J Pharmacogenomics* **4**, 371-381
210. Daub, H., Godl, K., Brehmer, D., Klebl, B., and Muller, G. (2004) Evaluation of kinase inhibitor selectivity by chemical proteomics. *Assay Drug Dev Technol* **2**, 215-224
211. Lolli, G., Thaler, F., Valsasina, B., Roletto, F., Knapp, S., Uggeri, M., Bachi, A., Matafora, V., Storici, P., Stewart, A., Kalisz, H. M., and Isacchi, A. (2003) Inhibitor affinity chromatography: profiling the specific reactivity of the proteome with immobilized molecules. *Proteomics* **3**, 1287-1298
212. Imanishi, S., and Harada, K. (2004) Proteomics approach on microcystin binding proteins in mouse liver for investigation of microcystin toxicity. *Toxicon* **43**, 651-659
213. Knockaert, M., Gray, N., Damiens, E., Chang, Y. T., Grellier, P., Grant, K., Fergusson, D., Mottram, J., Soete, M., Dubremetz, J. F., Le Roch, K., Doerig, C., Schultz, P., and Meijer, L. (2000) Intracellular targets of cyclin-dependent kinase inhibitors: identification by affinity chromatography using immobilised inhibitors. *Chem Biol* **7**, 411-422
214. Knockaert, M., and Meijer, L. (2002) Identifying in vivo targets of cyclin-dependent kinase inhibitors by affinity chromatography. *Biochem Pharmacol* **64**, 819-825
215. Knockaert, M., Wieking, K., Schmitt, S., Leost, M., Grant, K. M., Mottram, J. C., Kunick, C., and Meijer, L. (2002) Intracellular Targets of Paullones. Identification following affinity purification on immobilized inhibitor. *J Biol Chem* **277**, 25493-25501
216. Brehmer, D., Greff, Z., Godl, K., Blencke, S., Kurtenbach, A., Weber, M., Muller, S., Klebl, B., Cotten, M., Keri, G., Wissing, J., and Daub, H. (2005) Cellular targets of gefitinib. *Cancer Res* **65**, 379-382
217. Niture, S. K., Doneanu, C. E., Velu, C. S., Bailey, N. I., and Srivenugopal, K. S. (2005) Proteomic analysis of human O(6)-methylguanine-DNA methyltransferase by affinity chromatography and tandem mass spectrometry. *Biochem Biophys Res Commun*
218. Yaneva, M., and Tempst, P. (2003) Affinity capture of specific DNA-binding proteins for mass spectrometric identification. *Anal Chem* **75**, 6437-6448
219. Evans, M. J., Saghatelian, A., Sorensen, E. J., and Cravatt, B. F. (2005) Target discovery in small-molecule cell-based screens by in situ proteome reactivity profiling. *Nat Biotechnol* **23**, 1303-1307

Chapter 2

Double standards in quantitative proteomics: direct comparative assessment of difference in gel electrophoresis (DiGE) and metabolic stable isotope labeling

Eef H.C.Dirksen, Annemieke Kolkman*, Monique Slijper and Albert J.R.Heck*

*These authors contributed equally to this work

Department of Biomolecular Mass Spectrometry, Utrecht Institute for Pharmaceutical Sciences, Bijvoet Centre for Molecular Research, Utrecht University, Utrecht, The Netherlands

Based on: *Molecular and Cellular Proteomics*, **2005**, 4 (3), 255-266

ABSTRACT

Quantitative protein expression profiling is a crucial part of proteomics and requires methods that are able to efficiently provide accurate and reproducible differential expression values for proteins in two or more biological samples. Here, two state-of-the-art quantitative proteomics approaches, i.e. difference in gel electrophoresis (DiGE) and metabolic stable isotope labeling were evaluated in a direct comparative assessment. For that purpose, *Saccharomyces cerevisiae* was grown under well-defined experimental conditions in chemostat cultures under two single-nutrient limited growth conditions, using ^{14}N or ^{15}N ammonium sulfate as single nitrogen source. Following lysis and protein extraction from the two yeast samples, proteins were labeled using different fluorescent CyDyes. Subsequently, the yeast samples were mixed and the proteins separated by 2D gel electrophoresis.

Peptides resulting from the in-gel digestion of proteins were analyzed using MALDI-TOF mass spectrometry. Relative protein expression ratios between these two yeast samples were determined using both DiGE and metabolic stable isotope labeling. Focusing on a small, albeit representative set of proteins covering the whole gel-range, including some protein isoforms, and ranging from low to high abundance, it was observed that the correlation between these two methods of quantification is good: the differential ratios determined match the equation $R_{Met.Lab.} = 0.98\ R_{DiGE}$, with $r^2 = 0.89$. Although the correlation between DiGE and metabolic stable isotope labeling is exceptionally good, some advantages as well as disadvantages of both methods were observed. These are discussed and evaluated in relation to other (quantitative) approaches.

INTRODUCTION

The proteome is generally defined as the total protein complement of a genome present in cells and/or tissue (1). As a result of splice variation and/or post-translational modifications, the proteome is several orders of magnitude more complex than the genome. An extra factor that adds to the relative complexity of the proteome is the fact that protein abundance varies over time, either as a reaction to changes in the environment or during development. To understand these dynamic processes, which may lead to indications why *e.g.* "healthy" cells are different from "diseased" or "stressed" cells, or how cells change during differentiation, it is not only important to identify which proteins are involved but also to measure their differential expression levels. It is especially this latter notion that makes quantitative protein profiling an essential

part of proteomics, which requires technologies that accurately, reproducibly, and comprehensively quantify the protein content in biological samples (2-4).

Traditionally, and probably still, the most frequently used method to investigate differential protein abundances in large scale proteomics experiments on protein mixtures from cellular extracts or tissue is by two-dimensional gel electrophoresis (2D-GE) (5-9). In such experiments proteins are separated by their pI and molecular weight on a 2D gel and subsequently stained for visualization. The spot density on the gel is used to assess relative quantification through comparison with "matched" protein spots on parallel run 2D gels. For protein staining many protocols are in use (6), whereby Coomassie Brilliant Blue (CBB) and silver staining have found most widespread applications. These stains have appeared to be not ideal because of relatively poor detection sensitivity (CBB), or diminished peptide recovery from in-gel digested proteins for mass spectrometry (MS) (silver staining). Both CBB and silver staining also have a limited dynamic range. The accuracy of quantification depends on the intrinsic characteristics of the visualization methods. More recently, a variety of improvements and alternatives that are more reproducible and have an increased linear dynamic range have been introduced. Significant quality improvements have been achieved with the introduction of fluorescent stains like that of the SYPRO family (10), which besides an increase in linear dynamic range turned out to be satisfactorily compatible with MS analysis.

Another recently introduced novel approach in 2D gel based quantitative proteomics is the application of fluorescent cyanine dyes (Cy2, Cy3, Cy5) to label proteins before they are separated on a 2D gel (11-13). These fluorescent labels carry a *N*-hydroxysuccinimidyl ester functionality, designed to modify the ε-amino group of lysine residues in proteins. The three spectrally resolvable fluorophores carry a positive charge to compensate for the charge of the lysine that is lost upon labeling, thereby balancing the pI of the protein. The molecular masses of the CyDyes are approximately 450 Da, and will not significantly affect the protein migration in the second dimension. Taken together, the characteristics of these labels allow the analysis of up to three pools of protein samples simultaneously on a single 2D gel. This approach eliminates technical, i.e. gel-to-gel, variation which is the main limitation of 2D gel electrophoresis. In a standard protocol, two of the dyes (typically Cy3 and Cy5) are used to label two different pools of protein samples, while the third label (Cy2) is used to label an internal standard that consists of equal amounts of the two pools. This internal standard allows a correction for further experimental errors, thereby distinguishing biological from experimental variation (14). Since its introduction this so-called difference in gel electrophoresis (DiGE) approach has found

applications in quantitative proteomics (15-21), for instance in comparative quantitative proteomics of primitive hematopoietic cell populations (17).

In recent years also entirely different, mass spectrometry based, methods to assess protein expression levels have been developed. Herein, differential quantification is accomplished by labeling peptides/proteins with stable isotope tags. These techniques are quite different from radioactive isotope labeling (with ^{35}S methionine for instance) (22), which is probably still the most sensitive and accurate method to label/stain proteins, but which is rather hazardous. In stable isotope labeling, proteins or peptides in two sets of samples are differentially labeled using different stable isotope tags. These different isotope tags will produce specific mass-shifts in the mass spectra of peptides/proteins, which may then be used as internal standards in differential analysis. In this way, differential quantification by mass spectrometric analysis can be achieved. Many different stable isotope-containing labels have now been developed, which may be classified on the basis of *how* and *when* they are introduced into the protein samples. The stable isotope labels may be incorporated by chemical or biological means, at different stages of the proteomics experiment, i.e. from the start *in vivo* in cells or organisms up to the end by reacting the protein digest with appropriate labels just prior to mass spectrometric analysis (23,24).

Chemically, the stable isotope label can be incorporated via reactions with isotope-containing reagents at different functional groups in the peptides/proteins such as lysine side-chains or the free N-termini etc. (25-28). By using an isotope coded affinity tag, such as the biotinylated ICAT reagent that reacts selectively with free cysteines, stable isotope labeled peptides/proteins can be enriched prior to mass analysis (29-34). Alternatively, generation of C-terminal labeled peptides can be achieved by enzymatic digestion in heavy $H_2{}^{18}O$ water (35-40). More recently a novel chemical isotope labeling approach, named iTRAQ, was introduced that uses a multiplexed set of isobaric reagents that yield amine-derivatized peptides. The derivatized peptides are indistinguishable in MS, but exhibit intense low-mass MS/MS marker ions that may be used for relative quantification of proteins originating from up to four different sample pools (41). A disadvantage of these "chemical" approaches is that the stable isotope label is introduced into the sample only after several stages of sample preparation, such as cell lysis, protein extraction and/or even proteolysis. When the mixing of the differentially labeled samples occurs only after several of these sample preparation steps it is of ultimate importance in these approaches that the sample preparation is highly consistent.

Therefore, it is preferred to introduce the stable isotope label very early in the process. In such approaches, the cells or organisms need to be grown in defined media containing a stable isotope label that can be incorporated during protein synthesis (42-44). In a typical approach, termed metabolic labeling, growth medium is prepared in which a stable isotope labeled compound, like ^{15}N labeled ammonium sulfate, is used as the sole nitrogen source. Alternatively, also stable isotope labeled amino acids can be introduced into the medium, which will be incorporated (in the case of essential amino acids), during protein synthesis (45-49). So far metabolic labeling has been mostly applied to unicellular organisms, such as yeast (42), bacteria (43) and to tissue cell cultures (45) that can be easily grown on defined media in the laboratory. Recently, also the multicellular organisms *Caenorhabditis elegans* and *Drosophila melanogaster* (50) have been metabolically labeled, and lately this has even been extended to the isotope labeling of a complete rat (51) and potato plant (52).

In general, all these different quantitative proteomics approaches have their merits and limits and the method of choice often depends on the particular biological question. However, as far as accuracy and validation of methods in protein quantification is concerned, only very few reports exist in which different quantification techniques are directly compared (22,53,54). For instance, Lopez *et al*. (54) compared the quantification of about 400 protein spots stained by silver and SYPRO Ruby on 2D gels and found an overall correlation of just 0.75, with the largest deviation at lower protein abundances. Fievet *et al*. (22) compared protein quantities from yeast proteins labeled with radioactive ^{35}S or stained with CBB. They observed a very weak correlation and found the relative ratios determined by these two methods to vary for individual proteins from 0.37 to 1.86. As it is of absolute importance in quantitative proteomics that methods are able to accurately, reproducibly, and comprehensively quantify the protein content in biological samples, we set out to evaluate two current state-of-the-art quantitative approaches, difference in gel electrophoresis (DiGE) and metabolic stable isotope labeling, in a direct comparative assessment. For this, we used *Saccharomyces cerevisiae* that was grown under well-defined experimental conditions in chemostat cultures under two different single-nutrient limited growth conditions (i.e. nitrogen *versus* carbon) as a model system. One of the two yeast samples was grown in the chemostats using medium containing a stable isotope (i.e. ^{15}N) while the other was grown on natural isotope containing medium, with ammonium sulfate being the sole nitrogen source. Throughout this work the term ^{15}N indicates proteins, or peptides thereof, that were extracted from yeast grown on medium containing 98% $^{15}(NH_4)_2SO_4$ as sole nitrogen source, while proteins extracted from yeast grown on medium containing $(NH_4)_2SO_4$ are

referred to as the natural isotope. Following lysis and protein extraction, the two samples were fluorescently labeled using two different fluorescent CyDyes, prior to mixing. Proteins were separated by 2D gel electrophoresis and after in-gel digestion further analyzed by mass spectrometry. Protein expression levels of these two yeast samples were relatively quantified both using DiGE and metabolic stable isotope labeling. Focusing on a small, but representative set of protein spots with a wide variety in pI, M_r and abundance, we observe that, when excluding so-called on-off spots, the correlation between the two methods of quantification is very good, with the differential ratios determined following the equation $R_{Met.Lab.} = 0.98\ R_{DiGE}$, with a correlation coefficient r^2 of 0.89.

EXPERIMENTAL PROCEDURES

Strain and culture conditions

Wild-type *Saccharomyces cerevisiae* strain CEN.PK113-7D (*MAT***a**) (55) was grown at 30°C in 2-liter chemostats (Applikon), with a working volume of 1.0 liter as described in (56). Cultures were fed with a defined mineral medium that limited growth by either carbon or nitrogen with all other growth requirements in excess and at a constant residual concentration. The defined mineral medium composition was based on that described by Verduyn *et al* (57). The medium contained the following components: *Carbon-limited:* $(NH_4)_2SO_4$ 19 mM and glucose 42 mM; *Nitrogen-limited:* $(NH_4)_2SO_4$ 7.5 mM and glucose 330 mM.

Both cultures were started with $(^{14}NH_4)_2SO_4$ as sole nitrogen source (Merck, Darmstadt, Germany). In case the ^{15}N isotope was used, the medium vessel was replaced by a new vessel containing $(^{15}NH_4)_2SO_4$ (Isotec Inc, Miamisburg, USA) after 5 volume changes. The carbon-limited culture was fed with 98% $(^{15}NH_4)_2SO_4$ as supplied by Isotec. After five additional volume changes, a new steady-state was reached and samples for proteome analysis were taken. Dry-weight, metabolite, dissolved oxygen and gas profiles were constant over at least 3 volume changes prior to sampling. Samples dedicated to proteome analysis were sampled on ice and immediately centrifuged (5 min. at 0°C), washed twice with ice-cold sterile water and stored 5 times concentrated in water at -80°C.

Protein extraction

Protein extracts were prepared as described previously (58). Protein concentration was determined using the Plus One 2D Quant Kit (Amersham Biosciences). The protein samples were stored in aliquots at –80°C.

Labeling of proteins with CyDyes

Protein samples were prepared and labeled according to the manufacturers protocol. Briefly, 50 μg of protein was precipitated using the Plus One 2D Clean-Up Kit (Amersham Biosciences), dissolved in labeling buffer, and labeled at 0°C in the dark for 30 minutes with 400 pmoles of cyanine dye (Cy2, Cy3, Cy5; Amersham Biosciences), dissolved in 99.8% DMF (Sigma). The reaction was quenched by the addition of 1 μL of a 10 mM *L*-lysine solution (Merck) and left on ice for 10 minutes.

Two-dimensional gel electrophoresis

Two-dimensional gels were run as described before (58). Briefly, the three 50 μg aliquots of the Cy2, Cy3 and Cy5 labeled proteins were mixed and loaded on a 24-cm Immobiline Dry-Strip pH3-10 NL (Amersham Biosciences). Isoelectric focusing was carried out using an IPGphor (Amersham Biosciences) to a total of 50-55 kVh. After equilibration, strips were placed on top of 12.5% polyacrylamide gels and sealed with a solution of 1% (w/v) agarose containing a trace of bromophenol blue. Gels were run overnight at a constant power of 2 Watt until the bromophenol blue front had migrated to the bottom of the gel.

Image acquisition and analysis

Gels were scanned using the Typhoon 9400 Imager (Amersham Biosciences) according to the manufacturers protocol. Scans were acquired at 100 μm resolution. After cropping and filtering, images were subjected to automated Difference in-gel Analysis (DIA) and Biological Variation Analysis (BVA) using the Batch Processor of DeCyder software version 5.01 (Amersham Biosciences).

Post staining

2D gels were post-stained using silver staining as described by Shevchenko (59) with slight modifications. Briefly, after fixing and washing, the gels were sensitized using 0.04% sodium thiosulfate and impregnated with 0.1% silver nitrate at 4°C for 20 minutes. Development of the gel was performed using 3% sodium carbonate/0.05% formalin. Silver stained gels were scanned using a GS710 Calibrated Densitometer (Bio-Rad).

In-gel tryptic digestion

Protein spots of interest were digested in-gel with trypsin according to a slightly modified version of the protocol described by Wilm *et al.* (60). The gel pieces were first destained using 30 mM potasium ferricyanide and 100 mM

sodium thiosulfate solution, followed by washing and shrinking steps using 50 mM ammonium bicarbonate and acetonitrile, respectively. Proteins were digested overnight at 37^0C.

MALDI-MS and protein identification

Tryptic digests were desalted and concentrated with μC18-ZipTips (Millipore) and analysed on a Voyager DE-STR MALDI-TOF mass spectrometer (Applied Biosystems) using α-cyano-4-hydroxycinnamic acid as matrix. The MALDI-MS resolution for the peptides was typically ~10000. The raw MALDI-TOF spectra were processed using Data Explorer software (version 4.0, Applied Biosystems). The following process parameters were used before the final peak list was generated: advanced baseline correction, smoothing, and peak de-isotoping. The MALDI-MS spectra were internally calibrated using the singly protonated trypsin auto-digestion peaks at *m/z* 2273.159 and 2163.056. The MALDI-MS spectra were searched against the SwissProt database using a local Mascot search engine (61). The following settings were used: trypsin was used as enzyme, a maximum of 2 missed cleavages was allowed, the peptide tolerance was set at 150 ppm, carbamidomethyl cysteine and oxidized methionine were set as fixed and variable modification, respectively. The MALDI-MS spectra were searched twice against the SwissProt database. The first time using the above described parameters. The second time also by using the above described parameters with additional a newly defined fixed modification, assuming that all nitrogen atoms in the amino acids are ^{15}N labeled. In this way, both the natural abundance ^{14}N peptides as well as ^{15}N labeled peptides were identified, significantly increasing the confidence score for identification.

Protein expression ratio determination

Ratios of differentially expressed proteins (R_{DiGE}) were calculated using DeCyder (v5.01, Amersham Biosciences) for DiGE and show the fold change of the expression under nitrogen-limiting conditions *versus* carbon-limiting conditions (N/C). In the DeCyder output, an increase in protein abundance under nitrogen-limitation is expressed as a positive value (*e.g.* a two-fold increase = 2), while a decrease in protein abundance under nitrogen-limitation is expressed as a negative value (*e.g.* a two-fold decrease = -2).

For metabolic stable isotope labeling, proteins were relatively quantified as described previously (50). Briefly, peaks of all isotopes of the unlabeled peptide were integrated and divided by the integrated peak area of the ^{15}N labeled peptide. The integration was performed by zooming in on all the isotopes of the peptide of interest, and subsequently the area under all isotopes was

calculated in the Data Explorer software (version 4.0, Applied Biosystems) and was subsequently exported to Excel (Microsoft). The protein expression ratio nitrogen-limitation *versus* carbon-limitation was calculated for each peptide pair. This was performed for multiple peak pairs in the same MALDI-TOF mass spectrum and the $R_{Met.Lab.}$ was calculated as the average ratio of the multiple peak pairs. To enable a direct comparison with the DiGE quantitative data an increase in protein abundance under nitrogen-limitation was expressed as a positive value, while a decrease in protein abundance under nitrogen-limitation was expressed as a negative value. This representation is used throughout this work.

RESULTS

In this study, two state-of-the-art relative quantitative proteomics approaches, namely difference in gel electrophoresis (DiGE) and metabolic stable isotope labeling, were evaluated in a direct comparative assessment. To make an appropriate technological comparison while limiting any side effects, for instance due to experimental variation, the setup of the experiment is of vital importance. In Figure 1 and Table 1 the details of the selected experimental setup are shown.

Table 1. *Employed labeling strategy. Yeast sample composition of the 2D gels used for the direct comparative assessment. Yeast was grown under either nitrogen limited or carbon limited conditions in chemostat cultures. The growth media either contained natural isotope or ^{15}N labeled ammonium sulfate as the sole nitrogen source. Furthermore the CyDye labels, which are used to label the protein samples are indicated, i.e. Cy3 or Cy5.*

Gel	Limitation	Nitrogen source	CyDye
1	Nitrogen	Natural isotope	Cy3
	Carbon	^{15}N labeled	Cy5
2	Nitrogen	Natural isotope	Cy5
	Carbon	^{15}N labeled	Cy3

First of all, yeast cells were cultivated in well-controlled chemostats limited for either the carbon or nitrogen source. The only nitrogen source in these chemostat cultures was ammonium sulfate ($(NH_4)_2SO_4$). For the yeast cells grown under nitrogen limited conditions the natural isotope $(NH_4)_2SO_4$ was used, while in the carbon limited culture ^{15}N enriched $(NH_4)_2SO_4$ was used as sole

nitrogen source. Following cell lysis, proteins from these two chemostat yeast cultures were extracted. Complete incorporation of ^{15}N in yeast proteins was checked using tryptic digestion and mass spectrometry (data not shown). Subsequently, proteins were labeled with the fluorescent cyanine dyes. As described in Table 1, two DiGE experiments were performed: in the first experiment yeast grown under nitrogen limited conditions was labeled with Cy3 and ^{15}N labeled yeast grown under carbon limited conditions was labeled with Cy5. In the second DiGE experiment, the same chemostat cultured samples were used, but the two fluorescent dyes were reversed. Through this so-called dye swap an extra internal control for the DiGE experiments was incorporated. In addition, an internal standard that consisted of equal amounts of protein from both yeast samples and was labeled with Cy2, was used for the DiGE experiments. The DiGE internal standard was used to normalize the data, thereby limiting technical variation (14). Following fluorescent labeling, the samples were mixed and the proteins separated on one 2D gel. For the DiGE analysis, 2D gel images were acquired by fluorescence scanning and analyzed using DeCyder software. The DiGE analysis of the two separate fluorescent labeling experiments yielded two distinct values for the relative fold change in protein expression between the yeast grown under nitrogen-limited *versus* carbon-limited conditions. In theory these values should be equal, and any variation observed must be due to intrinsic variation in the DiGE methodology.

In order to allow analysis of the in-gel separated proteins, the gels were post-stained with silver. Upon comparison of the individual images, no differences concerning spot position were seen. Some differences between the staining methods were observed however, resulting in some spots being preferably 'stained' by either of the two (data not shown). Comparison of the linear dynamic range of DiGE and silver staining showed DiGE to have a range of at least four orders of magnitude, while that of silver staining had a maximum of two orders of magnitude, in agreement with previously reported data (54).

In this comparative evaluation a number of spots on the 2D gel (Figure 2) that ranged over the whole gel (with extensive variation in mass and pI) and varied over a wide range of protein expression levels (as determined by the DeCyder analysis) were picked.

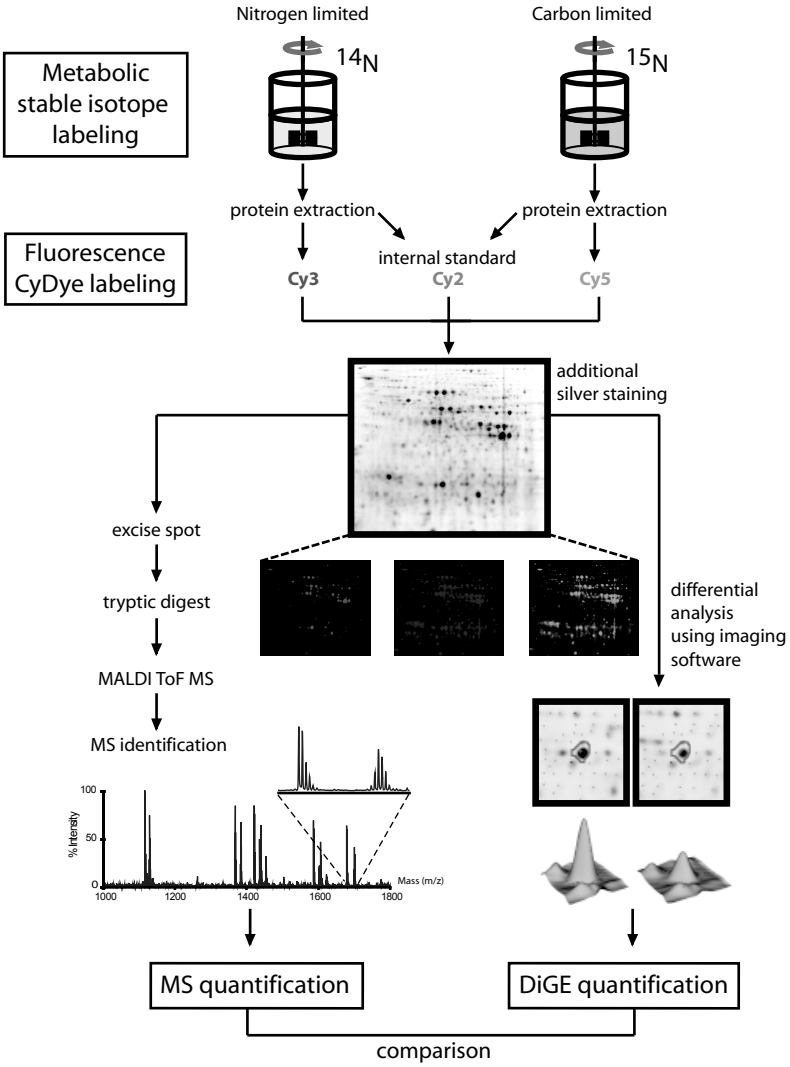


Figure 1. *Experimental setup for the direct comparative assessment of Difference in Gel Electrophoresis (DiGE) and metabolic stable isotope labeling. Saccharomyces cerevisiae was cultured in nitrogen-limited and carbon-limited chemostats in media containing either "natural" ammonium sulfate or ^{15}N labeled ammonium sulfate as the sole nitrogen source, respectively. Protein extracts from both cultures were prepared and proteins were labeled with the CyDyes as described in Table 1. A mixture of CyDye labeled proteins was prepared and run on a single 2D gel. 20 protein spots of interest were excised, digested in-gel with trypsin and analyzed with MALDI-TOF-MS to identify the protein. Protein expression ratios were determined in two ways; via comparing the fluorescence spot intensities (DiGE) and secondly via comparison of mass spectrometry peak areas of the unlabeled versus ^{15}N labeled peptide.*

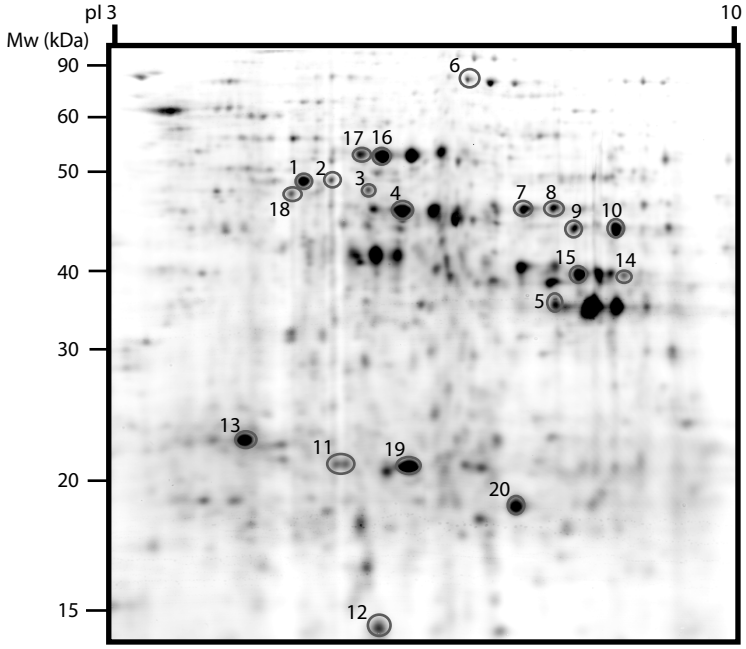


Figure 2. *Typical example of a 2D gel of a 1:1 mixture of protein extracts from chemostat-grown yeast cells, limited for either nitrogen or carbon. 150 µg of protein extract was separated in the first dimension on a 24 cm IPG strip (pI 3-10NL) and on a 12.5% SDS PAGE gel in the second dimension. Protein spots that were excised and analyzed by mass spectrometry are numbered and correspond to the numbers in Table 2.*

Table 2. *Protein expression ratios (nitrogen-limited versus carbon-limited) as determined by DiGE (R_{DiGE}) and stable isotope metabolic labeling ($R_{Met.Lab}$). In total 20 spots were relatively quantified on two 2D gels with both DiGE and stable isotope metabolic labeling. The spot numbering corresponds with the numbering in Figure 2. The protein identity, SwissProt accession number, the observed isoelectric point (pI) and approximate molecular weight (Mw), and the sequence coverage for each protein spot are indicated. Below R_{DiGE} column 1 and 2, the two ratios for the protein expression levels as obtained by the DiGE image analysis of the two separate gels are shown, followed by the average of the two values (Av) and the standard deviation (Std). Similar, in $R_{met.Lab}$ column 1 and 2 the two ratios (Av), as well as the inter protein standard deviation (Std.) and the number of peptides ion peak pairs (#) on which the relative quantification is based are indicated. In the last two columns the average ratio (Av.) and standard deviation (Std.) for the protein expression levels as obtained by the two metabolic stable isotope labeling experiment are given. A negative ratio indicates a down regulation of the protein in the nitrogen-limited culture, and a positive value an up regulation of the protein in the nitrogen-limited culture (Table can be found on page 53).*

Spot	Protein	SwissProt	pI	Mw (kDa)	Sequence coverage	R_{DiGE}				$R_{Met.Lab.}$							
						1	2	Av	Std	1			2			Av	Std
										Av	Std	#	Av	Std	#		
1	Gdh1p	P07262	5.56	49.5	30%	1.87	1.73	1.8	0.1	1.29	0.13	4	1.22	0.09	4	1.26	0.05
2	Gdh1p	P07262	5.58	49.5	38%	1.27	1.3	1.29	0.02	1.27	0.13	2	1.51	0.00	2	1.39	0.17
3	Gdh1p	P07262	5.62	49.5	40%	1.94	1.52	1.73	0.3	1.25	0.13	4	1.26	0.10	4	1.26	0.01
4	Eno2p	P00925	5.67	46.8	39%	1.62	1.49	1.56	0.09	1.07	0.34	4	1.4	0.27	4	1.24	0.23
5	Tdh3p	P00359	6.49	35.5	35%	1.26	1.6	1.43	0.24	1.11	0.06	4	1.22	0.07	4	1.17	0.08
6	Met6p	P05694	6.07	85.8	26%	1.97	1.8	1.89	0.12	1.83	0.15	2	1.3	0.04	2	1.57	0.37
7	Eno1p	P00924	6.17	46.5	54%	-1.27	-1.22	-1.25	0.04	-1.58	0.27	3	-3.57	0.39	3	-2.58	1.41
8	Eno1p	P00924	6.45	46.5	31%	-1.8	-1.73	-1.77	0.05	-1.83	0.21	3	-2.44	0.24	3	-2.14	0.43
9	Pgk1p	P00560	7.09	44.6	53%	1.01	1.07	1.04	0.04	1.33	0.26	3	1.16	0.45	3	1.25	0.12
10	Pgk1p	P00560	7.58	44.6	39%	1.17	1.33	1.25	0.11	1.02	0.03	4	1.01	0.05	4	1.01	0.01
11	Tpi1p	P00942	5.75	26.6	33%	1.6	1.46	1.53	0.1	1.85	0.35	4	1.5	0.19	4	1.68	0.25
12	Sod1p	P00445	5.63	15.5	71%	1.01	1.06	1.04	0.04	1.05	0.05	4	1.08	0.52	4	1.07	0.02
13	Hsp26p	P15992	5.31	24	43%	-5.99	-6.86	-6.43	0.62	-12.98	46.7	5	-34.44	12.5	5	-23.7	15.20
14	Adh2p	P00331	6.26	36.6	27%	-9.83	-13.3	-11.6	2.44	-9.65	52.3	3	-25.17	42.8	3	-17.4	11.00
15	Adh2p	P00331	7.15	36.6	17%	-14.4	-41.4	-27.9	19.2	-23.86	16.7	3	-37.97	12.5	3	-30.9	9.98
16	Pdc1p	P06169	5.8	61.4	35%	2.75	2.45	2.6	0.21	4.23	3.23	4	4.99	1.40	4	4.61	0.54
17	Pdc1p	P06169	5.6	61.4	34%	3.46	2.77	3.12	0.49	10.30	5.30	4	20.36	20.14	4	15.3	7.11
18	Arg1p	P22768	5.53	47	28%	1.4	1.22	1.31	0.13	1.30	0.17	4	1.44	0.24	3	1.37	0.10
19	Tpi1p	P00942	5.75	26.6	28%	1.75	1.76	1.75	0.02	1.51	0.21	5	1.54	0.16	5	1.53	0.04
20	Tdh3p	P00359	6.49	18	24%	-1.07	-1.12	-1.09	0.05	-1.22	0.15	2	-1.31	0.10	2	-1.26	0.05

These protein spots were excised, and after tryptic digestion, the proteins were identified by peptide mass fingerprinting using a MALDI-TOF mass spectrometer. Out of the spots analyzed, 20 spots were selected that originated exclusively from a single protein, as revealed by mass spectrometric analysis, i.e. all peptides observed in the mass spectra originated from that protein, excluding the possibility that the spot intensity as measured by DiGE originated from more than one protein. The measured ratios of protein expression of these 20 selected proteins obtained by using DiGE are given in Table 2. Protein identifications of these 20 protein spots are given in Table 2, and revealed that the twenty spots corresponded to 12 different proteins, with 7 proteins appearing in more than one spot on the 2D gel, indicating the presence of protein isoforms and/or post-translational modifications. Peptides carrying a fluorophore modification on a lysine residue were not observed in our MALDI-TOF spectra, which was expected as only a very small percentage (< 3%) of the proteins is labeled with the CyDyes.

As each selected spot contains proteins from the isotope labeled and unlabeled cultures the resulting MALDI peptide fingerprint mass spectra displayed numerous peptide pairs (Figure 1). Differential quantification was achieved by comparing the peak areas of the natural isotope containing and ^{15}N labeled peptides, averaged over all peptide pairs available. Thus, using this metabolic stable isotope labeling approach, the relative protein expression levels between the yeast samples grown under nitrogen-limited *versus* carbon-limited conditions were determined in duplicate. The protein expression ratios measured by stable isotope labeling, the number of peptides pairs on which quantification is based, as well as the intra-protein quantification standard deviation are also given in Table 2.

In Figure 3, we have focused on a few spot/protein examples, showing that both with DiGE and stable isotope labeling a wide range of differential expression ratios can be determined. 3D views of the fluorescent abundance of a protein spot in yeast grown under nitrogen-limitation (Cy3) and carbon-limitation (Cy5) next to a typical peptide ion peak pair measured in the MALDI-TOF spectra of the same protein spot are given. In Figure 3 the results for two proteins that appeared in multiple spots on the gel, i.e. Adh2p and Pdc1p are also depicted.

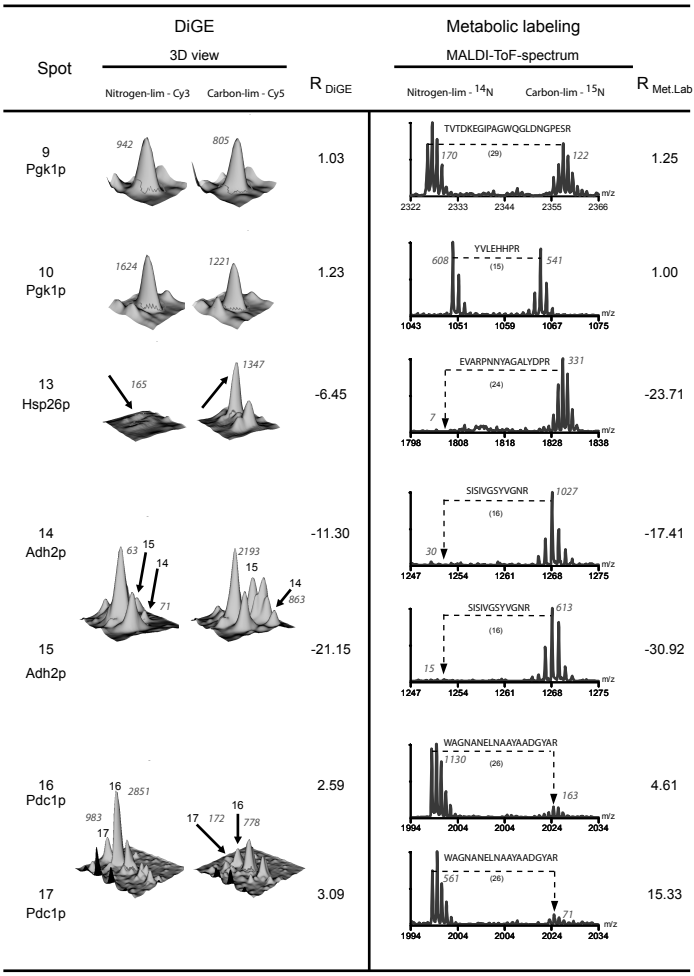


Figure 3. *Typical examples of relative protein quantification by DiGE and metabolic stable isotope labeling. On the left, next to the column with the spot number and protein name, 3D views of the fluorescent intensities of the spots are shown. In the given example, protein spot intensities in the nitrogen-limited chemostat culture were visualized using Cy3 fluorescence and protein spot intensities in the carbon-limited chemostat culture were visualized using Cy5 fluorescence. On the right, a typical tryptic peptide ion pair of the same protein spot measured by MALDI-TOF-MS is shown. The sequence of this peptide is indicated and in parentheses the number of nitrogen atoms of that particular peptide is indicated. Protein ratios, as defined in the experimental section, obtained by both methods (R_{DiGE} and $R_{Met.Lab}$) are indicated next to the 3D views and the peptide ion pairs, respectively. For both quantification methods, the signal to noise ratio for the individual protein signals (DiGE) and peptide signals (metabolic stable isotope labeling) are given in italic script revealing that for low abundant proteins the S/N ratios are generally smaller in the mass spectrometric approach.*

DISCUSSION

High-throughput proteome analyses, in which thousands of proteins are analyzed in a single experiment, call for design standards and guidelines that enable proper validation of the quality of the proteome data set and the conclusions drawn from the results. Recently, two reports appeared that address these issues and suggest parameters to judge the quality both of peptide and protein identification data (62) and of 2D gel based proteomics analyses (63). These are important steps towards quality control and validation. A next, equally essential, step is the cross-comparison and validation of different quantification techniques that are used in proteomics. The aim of this report is to contribute to this process and therefore two currently popular quantification methods were assessed in a direct comparative approach. The results of this comparison are comprehensively given in Table 2, with some illustrative detailed experimental results depicted in Figure 3.

Comparison of DiGE versus metabolic stable isotope labeling

A more direct comparison of both DiGE and stable isotope labeling methods for quantification of proteins is shown in Figure 4. Performing both the DiGE and metabolic stable isotope labeling quantification in duplicate provides a measure for the experimental standard deviation in the quantification by both methods. For both DiGE and the metabolic stable isotope labeling experiments the average ratio of the two separate measurements was determined, which is given in Table 2. These average ratios were used for a comparison between the DiGE and stable isotope labeling based quantification. Therefore, we divided the average ratio determined by metabolic stable isotope labeling by the ratio determined by DiGE for all 20 proteins. Theoretically, when both methods would provide accurate quantitative results, these values should be 1 for all individual protein spots. These divided ratios of the 20 spots are plotted in Figure 4A sorted by descending fold change values. Inspecting Figure 4A it is clear that the ratio between the fold changes observed by DiGE and metabolic stable isotope labeling are indeed close to 1, in particular when the ratio in protein expression between yeast grown under nitrogen-limited *versus* carbon-limited conditions is between -3 and 3.

When the average ratio determined by DiGE *versus* the one measured by stable isotope labeling is plotted, as shown in Figure 4B, a good correlation is found for this limited set of data (taking 15 out of the 20 spots). The data could be fitted with a linear relationship between the two determined ratios following the equation $R_{Met.Lab.} = 0.98\ R_{DiGE}$, with a r^2-value of 0.89, whereby $R_{Met.Lab.}$ is the average ratio determined by metabolic stable isotope labeling and $R_{DiGE.}$ is the

average ratio determined by DiGE. The obtained coefficient of one and the correlation r^2-value of 0.89 indicate that the differential quantification by metabolic stable isotope labeling and DiGE are within the margin of error equivalent.

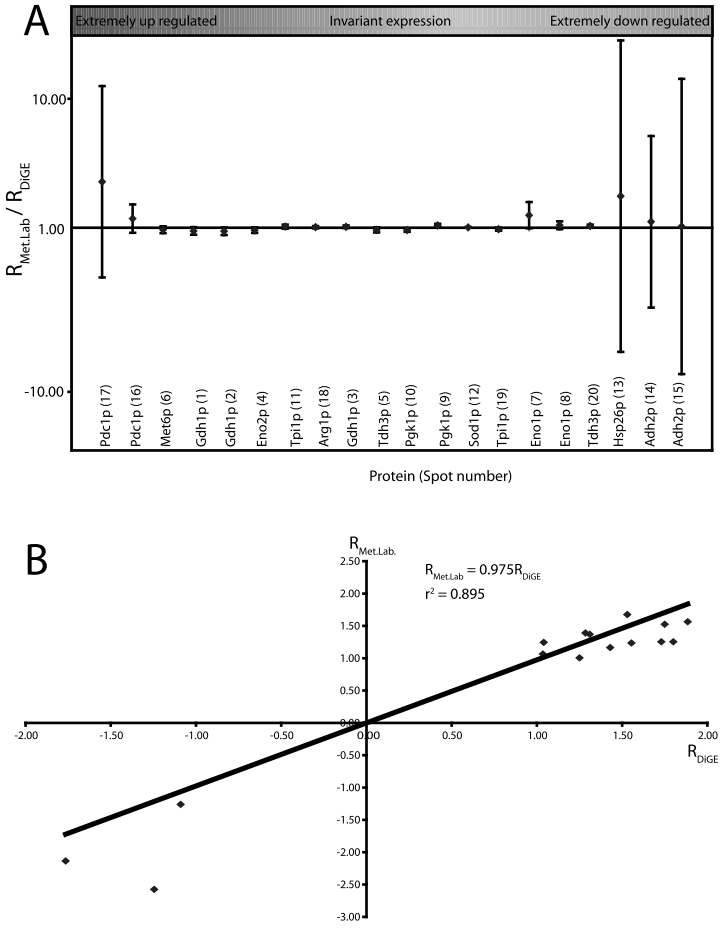


Figure 4. *Correlation between the average ratios obtained by DiGE (R_{DiGE}) and metabolic labeling ($R_{Met.Lab.}$). (A) The ratios obtained by metabolic stable isotope labeling are divided by the ratios obtained from the DiGE analysis ($R_{Met.Lab.}/R_{DiGE}$) and centered around 1, the value expected for a optimal correlation. (B) Correlation plot of the ratios obtained by DiGE (x-axis) against the ratios obtained via metabolic stable isotope labeling (y-axis). Extreme ratio's (on-off spots), that show large standard deviations, were excluded from this graph, leaving only ratios between -3 and 3.*

The five spots outside the limited range described above are the "on-off" spots that show "extreme" fold changes in protein expression. For these particular spots it was observed that the calculated R is primarily determined by the low signal-to-noise (S/N) values of the "off"-spot of either the fluorescent intensities on the DiGE gels or the peptide ion signals. Our data indicate that the S/N values in the mass spectrometric approach are generally smaller than in DiGE approach, especially for the low abundant proteins (see Figure 3) and this value also differs between individual spots on the MALDI target, depending on composition and crystallization quality. Therefore the ratio R is more difficult to accurately determine in the mass spectrometric approach. These smaller S/N values lead also to larger standard deviations for the calculated protein ratios R in the metabolic stable isotope labeling approach compared to fluorescent labeling (see Table 2). Overall, this demonstrates that DiGE may be better suited to determine changes of low abundant proteins and proteins that show extreme changes in expression.

Stable isotope labeling has experienced a dramatic increase in popularity in recent years in quantitative proteomics applications (43,44,50,64-68), thereby replacing, to some extent, conventional 2D gel-based approaches. This is probably due to the fact that at present stable isotope labeling is considered to be one of the most accurate ways to relatively quantify protein expression levels and additionally stable isotope labeling may be combined with (multidimensional) LC MS/MS approaches. As described in the introduction, in stable isotope labeling there are quite a few alternative approaches to introduce the label, either by chemical introduction of the isotope label (*e.g.* ICAT (29), iTRAQ (41)) or biological introduction of the label (^{15}N or ^{13}C metabolic labeling (50), SILAC (45,47)). The advantages and disadvantages of the different stable isotope labeling approaches have been discussed in detail in several reviews (23,44,66). Here, we just restate that some of the major advantages of the metabolic stable isotope labeling approach chosen here are that the label used for quantification is introduced very early on in the procedure (during cell growth), thereby decreasing the potential effect of differential losses in subsequent steps during sample preparation and additionally all proteins, and even all peptides, are uniformly labeled, increasing the probability that proteins may be quantitated by a larger set of peptide pairs, which is essential for accurate quantification. The latter is a major limitation in the stable isotope labeling approach, whereby in most reported experiments so far, the quantification of proteins is often only based on a single or just a few peptides per protein, hampering a meaningful error analysis in the quantification.

This study, and other DiGE experiments, reveal that with the implementation of pre-separation fluorescent dyes for protein labeling an alternative method capable of determining both small and large changes in protein expression has been added to the quantitative proteomics toolbox, producing accurate differential expression data. Compared to more conventional staining methods used in 2D gel electrophoresis, DiGE has a large dynamic range, allowing both the differential analysis of high and low abundant proteins. In the differential analysis of individual proteins, DiGE is probably even better than the stable isotope labeling approach, whereby the S/N level in the latter is largely dependent on the sensitivity and accuracy of the mass spectrometer used, and the complexity of the sample analyzed. Concerning the sensitivity and limits of stable isotope labeling and DiGE, it was observed that low abundant protein spots that still could be detected and quantified using DiGE, could not be detected and consequently not quantified using the mass spectrometry-based approach.

Another clear advantage of DiGE over metabolic stable isotope labeling is the general applicability: all protein samples irrespective of their origin (*e.g.* clinical samples) can be labeled as long as they contain lysine residues. In that sense, metabolic stable isotope labeling is limited to more simple uni- and multi-cellular organisms. Chemical introduction of stable isotopes such as in ICAT and iTRAQ on the other hand, are not hampered by this limitation.

The fact that proteins quantified on the gel still need to be identified, thereby requiring subsequent mass spectrometric analysis, might be considered as a disadvantage of the DiGE technology. Additionally, as with every 2D gel based technology, only subsequent analysis of protein spots, by for instance mass spectrometry, can reveal whether the spot of interest is "pure", i.e. originating from just one protein. If two or more proteins do co-migrate on the gel relative quantification is impossible. However, in general this report shows that the quantifications by metabolic stable isotope labeling and DiGE are in very good agreement.

Interestingly, the combined approach of stable isotope labeling and DiGE has, besides the achieved twofold quantification/validation, some other unique advantages, particularly in that some of the disadvantages of each of the methods are compensated by the other.

The illustrative examples in Figure 3 not only show that both methods provide similar results in up- and down-regulation, but also directly point out some intrinsic advantages of the combined 2D gel and stable isotope labeling approach used here. For instance, spot number 14 and 15, both identified as Adh2p, differ only in pI, thereby indicating that these proteins are most likely isoforms and/or post-translationally modified. Both forms are extremely up

regulated under carbon-limiting conditions. Also spot 16 and 17 are identified as "identical" proteins, i.e. Pdc1p, and differ only in pI. Interestingly, although both these Pdc1p isoforms are significantly down regulated under carbon-limiting conditions, both DiGE and stable isotope labeling indicate that spot 17 is more down regulated than spot 16 (Table 2 and Figure 3). In particular these data reveal an advantage of using 2D gel approaches instead of the direct analysis of total cell lysate digests by for instance a combination of stable isotope labeling and multidimensional LC. In this latter approach the peptide WAGNANELNAAYAADGYAR that was used for quantification of spot 16 and 17 (see Figure 3) would be analyzed only once and a single ratio would be determined averaged over the different protein isoforms, leading to erroneous quantification.

Another strong advantage of the combined approach of stable isotope labeling and DiGE is that on the one hand the mass spectrometric analysis can be used to ensure that the spot on the gel originates from only one protein, excluding co-migrating proteins in the analysis, on the other hand the DiGE separation provides directly multiple peptide pairs, all belonging to the same isoform of the protein, which facilitates quantitative analysis.

Comparison of gel-based versus non-gel based technologies

Although the combined stable isotope labeling and DiGE approach has advantages certain disadvantages linked to both methods remain. Classical 2D gel electrophoresis is labor intensive, hard-to-automate and is sensitive to technical variation. Additionally, despite the high-resolution separation capabilities of the 2D gel approach, certain classes of proteins (*e.g.* hydrophobic proteins, or those with high molecular weights and/or extreme pI values) are normally underrepresented in these analyses and moreover, the risk of overlapping proteins is introduced, hampering quantification of the individual proteins.

Alternative separation approaches based on liquid chromatographic separation of peptides resulting from proteolytically digested proteins from complete lysates, such as for example in MudPIT (69,70) were originally thought to replace 2D gel-based approaches, as they generally lead to higher-throughput and wider coverage of the full proteome. However, the direct LC based approaches have also their own intrinsic disadvantages, as they are more difficult to use for quantitative proteomics and for the analysis of protein isoforms. The identification and quantification of post-translationally modified proteins or protein isoforms in direct LC based approaches, for example, is only possible when the actual modified peptide is detected, significantly reducing the chance of quantification of the different protein forms. With all these pro- and cons of

the gel- and non-gel-based approaches it is becoming increasingly clear that both LC and gel-based technologies are more or less complementary not only in protein identification, but particularly in protein quantification.

In summary, both gel-based and liquid chromatography-based methods have their advantages and remaining challenges in quantitative proteomics. The metabolic stable isotope labeling and DiGE approach, comparatively assessed here, are both able to provide efficiently accurate and reproducible differential expression values for proteins in two or more biological samples and may therefore find wide applications in proteomics research. Combining the two methods not only allowed a direct validation of the two methods, but also revealed unique strong features, particularly in that some of the disadvantages of each of the methods could be compensated by the other.

ACKNOWLEDGEMENTS

We thank Pascale Daran-Lapujade and Jack Pronk (Delft University) for providing the metabolically stable isotope labeled, chemostat cultured, yeast samples. We thank Joost Gouw for making it possible to perform a database search with ^{15}N labeled peptides in the Mascot software. AK acknowledges financial support by DSM (The Netherlands). This work was supported by the Netherlands Proteomics Centre (http://www.netherlandsproteomicscentre.nl/).

REFERENCES

1. Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J. C., Yan, J. X., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L., and Hochstrasser, D. F. (1996) From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)* **14**, 61-65
2. Smith, R. D. (2000) Probing proteomes--seeing the whole picture? *Nat Biotechnol* **18**, 1041-1042
3. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**, 198-207
4. Hamdan, M., and Righetti, P. G. (2002) Modern strategies for protein quantification in proteome analysis: advantages and limitations. *Mass Spectrom. Rev.* **21**, 287-302
5. O'Farrell, P. H. (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**, 4007-4021
6. Righetti, P. G., Campostrini, N., Pascali, J., Hamdan, M., and Astner, H. (2004) Quantitative proteomics: a review of different methodologies. *Eur J Mass Spectrom (Chichester, Eng)* **10**, 335-348
7. Gorg, A., Obermaier, C., Boguth, G., Harder, A., Scheibe, B., Wildgruber, R., and Weiss, W. (2000) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* **21**, 1037-1053
8. Rabilloud, T. (2002) Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. *Proteomics* **2**, 3-10
9. Lilley, K. S., Razzaq, A., and Dupree, P. (2002) Two-dimensional gel electrophoresis: recent advances in sample preparation, detection and quantitation. *Curr Opin Chem Biol* **6**, 46-50
10. Yan, J. X., Harry, R. A., Spibey, C., and Dunn, M. J. (2000) Postelectrophoretic staining of proteins separated by two-dimensional gel electrophoresis using SYPRO dyes. *Electrophoresis* **21**, 3657-3665
11. 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
12. Unlu, M. (1999) Difference gel electrophoresis. *Biochem. Soc. Trans.* **27**, 547-549
13. Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., Pognan, F., Hawkins, E., Currie, I., and Davison, M. (2001) Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* **1**, 377-396
14. Alban, A., David, S. O., Bjorkesten, L., Andersson, C., Sloge, E., Lewis, S., and Currie, I. (2003) A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics* **3**, 36-44
15. Zhou, G., Li, H., DeCamp, D., Chen, S., Shu, H., Gong, Y., Flaig, M., Gillespie, J. W., Hu, N., Taylor, P. R., Emmert-Buck, M. R., Liotta, L. A., Petricoin, E. F., 3rd, and Zhao, Y. (2002) 2D differential in-gel electrophoresis for the identification of esophageal scans cell cancer-specific protein markers. *Mol Cell Proteomics* **1**, 117-124
16. Hu, Y., Wang, G., Chen, G. Y. J., Fu, X., and Yao, S. Q. (2003) Proteome analysis of *Saccharomyces cerevisiae* under metal stress by two-dimensional differential gel electrophoresis. *Electrophoresis* **24**, 1458-1470
17. Evans, C. A., Tonge, R., Blinco, D., Pierce, A., Shaw, J., Lu, Y., Hamzah, H. G., Gray, A., Downes, C. P., Gaskell, S. J., Spooncer, E., and Whetton, A. D. (2004) Comparative proteomics of primitive hematopoietic cell populations reveals differences in expression of proteins regulating motility. *Blood* **103**, 3751-3759
18. Wang, D., Jensen, R., Gendeh, G., Williams, K., and Pallavicini, M. G. (2004) Proteome and transcriptome analysis of retinoic acid-induced differentiation of human acute promyelocytic leukemia cells, NB4. *J Proteome Res* **3**, 627-635
19. Tian, Q., Stepaniants, S. B., Mao, M., Weng, L., Feetham, M. C., Doyle, M. J., Yi, E. C., Dai, H., Thorsson, V., Eng, J., Goodlett, D., Berger, J. P., Gunter, B., Linseley, P. S., Stoughton, R. B., Aebersold, R., Collins, S. J., Hanlon, W. A., and Hood, L. E. (2004)

Integrated Genomic and Proteomic Analyses of Gene Expression in Mammalian Cells. *Mol Cell Proteomics* **3**, 960-969
20. Kleno, T. G., Leonardsen, L. R., Kjeldal, H. O., Laursen, S. M., Jensen, O. N., and Baunsgaard, D. (2004) Mechanisms of hydrazine toxicity in rat liver investigated by proteomics and multivariate data analysis. *Proteomics* **4**, 868-880
21. Friedman, D. B., Hill, S., Keller, J. W., Merchant, N. B., Levy, S. E., Coffey, R. J., and Caprioli, R. M. (2004) Proteome analysis of human colon cancer by two-dimensional difference gel electrophoresis and mass spectrometry. *Proteomics* **4**, 793-811
22. Fievet, J., Dillmann, C., Lagniel, G., Davanture, M., Negroni, L., Labarre, J., and De Vienne, D. (2004) Assessing factors for reliable quantitative proteomics based on two-dimensional gel electrophoresis. *Proteomics* **4**, 1939-1949
23. Romijn, E. P., Krijgsveld, J., and Heck, A. J. (2003) Recent liquid chromatographic-(tandem) mass spectrometric applications in proteomics. *J Chromatogr A* **1000**, 589-608
24. Julka, S., and Regnier, F. (2004) Quantification in proteomics through stable isotope coding: a review. *J Proteome Res* **3**, 350-363
25. Geng, M., Ji, J., and Regnier, F. E. (2000) Signature-peptide approach to detecting proteins in complex mixtures. *J Chromatogr A* **870**, 295-313
26. Ji, J., Chakraborty, A., Geng, M., Zhang, X., Amini, A., Bina, M., and Regnier, F. (2000) Strategy for qualitative and quantitative analysis in proteomics based on signature peptides. *J Chromatogr B Biomed Sci Appl* **745**, 197-210
27. Munchbach, M., Quadroni, M., Miotto, G., and James, P. (2000) Quantitation and facilitated de novo sequencing of proteins by isotopic N-terminal labeling of peptides with a fragmentation-directing moiety. *Anal Chem* **72**, 4047-4057
28. Goodlett, D. R., Keller, A., Watts, J. D., Newitt, R., Yi, E. C., Purvine, S., Eng, J. K., von Haller, P., Aebersold, R., and Kolker, E. (2001) Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation. *Rapid Commun. Mass Spectrom.* **15**, 1214-1221
29. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**, 994-999
30. Han, D. K., Eng, J., Zhou, H., and Aebersold, R. (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat Biotechnol* **19**, 946-951
31. Griffin, T. J., Han, D. K., Gygi, S. P., Rist, B., Lee, H., Aebersold, R., and Parker, K. C. (2001) Toward a high-throughput approach to quantitative proteomic analysis: expression-dependent protein identification by mass spectrometry. *J. Am. Soc. Mass Spectrom.* **12**, 1238-1246
32. Zhou, H., Ranish, J. A., Watts, J. D., and Aebersold, R. (2002) Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. *Nat Biotechnol* **20**, 512-515
33. Smolka, M. B., Zhou, H., Purkayastha, S., and Aebersold, R. (2001) Optimization of the isotope-coded affinity tag-labeling procedure for quantitative proteome analysis. *Anal. Biochem.* **297**, 25-31
34. Smolka, M., Zhou, H., and Aebersold, R. (2002) Quantitative protein profiling using two-dimensional gel electrophoresis, isotope-coded affinity tag labeling, and mass spectrometry. *Mol Cell Proteomics* **1**, 19-29
35. Schnolzer, M., Jedrzejewski, P., and Lehmann, W. D. (1996) Protease-catalyzed incorporation of ^{18}O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. *Electrophoresis* **17**, 945-953
36. Mirgorodskaya, O. A., Kozmin, Y. P., Titov, M. I., Korner, R., Sonksen, C. P., and Roepstorff, P. (2000) Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using ^{18}O-labeled internal standards. *Rapid Commun. Mass Spectrom.* **14**, 1226-1232

37. Wang, Y. K., Ma, Z., Quinn, D. F., and Fu, E. W. (2001) Inverse ^{18}O labeling mass spectrometry for the rapid identification of marker/target proteins. *Anal Chem* **73**, 3742-3750
38. Stewart, II, Thomson, T., and Figeys, D. (2001) ^{18}O labeling: a tool for proteomics. *Rapid Commun. Mass Spectrom.* **15**, 2456-2465
39. Yao, X., Freas, A., Ramirez, J., Demirev, P. A., and Fenselau, C. (2001) Proteolytic ^{18}O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal Chem* **73**, 2836-2842
40. Reynolds, K. J., Yao, X., and Fenselau, C. (2002) Proteolytic ^{18}O labeling for comparative proteomics: evaluation of endoprotease Glu-C as the catalytic agent. *J Proteome Res* **1**, 27-33
41. Hardt, M., Witkowska, H. E., Webb, S., Thomas, L. R., Dixon, S. E., Hall, S. C., and Fisher, S. J. (2005) Assessing the effects of diurnal variation on the composition of human parotid saliva: quantitative analysis of native peptides using iTRAQ reagents. *Anal Chem* **77**, 4947-4954
42. Oda, Y., Huang, K., Cross, F. R., Cowburn, D., and Chait, B. T. (1999) Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci U S A* **96**, 6591-6596
43. Conrads, T. P., Alving, K., Veenstra, T. D., Belov, M. E., Anderson, G. A., Anderson, D. J., Lipton, M. S., Pasa-Tolic, L., Udseth, H. R., Chrisler, W. B., Thrall, B. D., and Smith, R. D. (2001) Quantitative analysis of bacterial and mammalian proteomes using a combination of cysteine affinity tags and ^{15}N-metabolic labeling. *Anal Chem* **73**, 2132-2139
44. Goshe, M. B., and Smith, R. D. (2003) Stable isotope-coded proteomic mass spectrometry. *Curr. Opin. Biotechnol.* **14**, 101-109
45. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376-386
46. Ong, S. E., Kratchmarova, I., and Mann, M. (2003) Properties of ^{13}C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J Proteome Res* **2**, 173-181
47. Everley, P. A., Krijgsveld, J., Zetter, B. R., and Gygi, S. P. (2004) Quantitative cancer proteomics: stable isotope labeling with amino acids in cell culture (SILAC) as a tool for prostate cancer research. *Mol Cell Proteomics* **3**, 729-735
48. Ibarrola, N., Kalume, D. E., Gronborg, M., Iwahori, A., and Pandey, A. (2003) A proteomic approach for quantitation of phosphorylation using stable isotope labeling in cell culture. *Anal Chem* **75**, 6043-6049
49. Blagoev, B., Kratchmarova, I., Ong, S. E., Nielsen, M., Foster, L. J., and Mann, M. (2003) A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nat Biotechnol* **21**, 315-318
50. Krijgsveld, J., Ketting, R. F., Mahmoudi, T., Johansen, J., Artal-Sanz, M., Verrijzer, C. P., Plasterk, R. H. A., and Heck, A. J. R. (2003) Metabolic labeling of *C-elegans* and *D-melanogaster* for quantitative proteomics. *Nat Biotechnol* **21**, 927-931
51. Wu, C. C., MacCoss, M. J., Howell, K. E., Matthews, D. E., and Yates, J. R., 3rd (2004) Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal Chem* **76**, 4951-4959
52. Ippel, J. H., Pouvreau, L., Kroef, T., Gruppen, H., Versteeg, G., van den Putten, P., Struik, P. C., and van Mierlo, C. P. (2004) In vivo uniform ^{15}N-isotope labelling of plants: using the greenhouse for structural proteomics. *Proteomics* **4**, 226-234
53. Rabilloud, T., Strub, J. M., Luche, S., van Dorsselaer, A., and Lunardi, J. (2001) A comparison between Sypro Ruby and ruthenium II tris (bathophenanthroline disulfonate) as fluorescent stains for protein detection in gels. *Proteomics* **1**, 699-704
54. Lopez, M. F., Berggren, K., Chernokalskaya, E., Lazarev, A., Robinson, M., and Patton, W. F. (2000) A comparison of silver stain and SYPRO Ruby Protein Gel Stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis* **21**, 3673-3683

55. van Dijken, J. P., Bauer, J., Brambilla, L., Duboc, P., Francois, J. M., Gancedo, C., Giuseppin, M. L., Heijnen, J. J., Hoare, M., Lange, H. C., Madden, E. A., Niederberger, P., Nielsen, J., Parrou, J. L., Petit, T., Porro, D., Reuss, M., van Riel, N., Rizzi, M., Steensma, H. Y., Verrips, C. T., Vindelov, J., and Pronk, J. T. (2000) An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme Microb. Technol.* **26**, 706-714
56. Boer, V. M., de Winde, J. H., Pronk, J. T., and Piper, M. D. (2003) The genome-wide transcriptional responses of *Saccharomyces cerevisiae* grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur. *J. Biol. Chem.* **278**, 3265-3274
57. Verduyn, C., Postma, E., Scheffers, W. A., and Van Dijken, J. P. (1992) Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* **8**, 501-517
58. Kolkman, A., Olsthoorn, M. M., Heeremans, C. E., Heck, A. J., and Slijper, M. (2005) Comparative proteome analysis of Saccharomyces cerevisiae grown in chemostat cultures limited for glucose or ethanol. *Mol Cell Proteomics* **4**, 1-11
59. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* **68**, 850-858
60. Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996) Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**, 466-469
61. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567
62. Carr, S., Aebersold, R., Baldwin, M., Burlingame, A., Clauser, K., and Nesvizhskii, A. (2004) The need for guidelines in publication of peptide and protein identification data: Working Group on Publication Guidelines for Peptide and Protein Identification Data. *Mol Cell Proteomics* **3**, 531-533
63. Celis, J. E. (2004) Gel-based Proteomics: What Does MCP Expect? *Mol Cell Proteomics* **3**, 949
64. Washburn, M. P., Ulaszek, R., Deciu, C., Schieltz, D. M., and Yates, J. R., 3rd (2002) Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Anal Chem* **74**, 1650-1657
65. Washburn, M. P., Ulaszek, R. R., and Yates, J. R., 3rd (2003) Reproducibility of quantitative proteomic analyses of complex biological mixtures by multidimensional protein identification technology. *Anal Chem* **75**, 5054-5061
66. Ong, S. E., Foster, L. J., and Mann, M. (2003) Mass spectrometric-based approaches in quantitative proteomics. *Methods* **29**, 124-130
67. Brand, M., Ranish, J. A., Kummer, N. T., Hamilton, J., Igarashi, K., Francastel, C., Chi, T. H., Crabtree, G. R., Aebersold, R., and Groudine, M. (2004) Dynamic changes in transcription factor complexes during erythroid differentiation revealed by quantitative proteomics. *Nat Struct Mol Biol* **11**, 73-80
68. Zhou, H., Boyle, R., and Aebersold, R. (2004) Quantitative protein analysis by solid phase isotope tagging and mass spectrometry. *Methods Mol Biol* **261**, 511-518
69. Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., and Yates, J. R., 3rd (1999) Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* **17**, 676-682
70. Washburn, M. P., Wolters, D., and Yates, J. R., 3rd (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**, 242-247

Chapter 3

Human lymphoblastoid proteome analysis reveals role for the INHAT complex in the DNA double strand break response

Eef H.C. Dirksen[1], Jacqueline Cloos[2], Boudewijn J.M. Braakhuis[2], Ruud H. Brakenhoff[2], Albert J.R. Heck[1], Monique Slijper[1]

[1] Department of Biomolecular Mass Spectrometry, Utrecht Institute for Pharmaceutical Sciences, Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands
[2] Section Tumor Biology, Department of Otolaryngology/Head-Neck Surgery, VU University Medical Center, Amsterdam, The Netherlands.

Based on: *Cancer Research*, **2006**, 66 (3), 1473-1480

ABSTRACT

A DNA double strand break (DSB) is highly cytotoxic, it emerges as the type of DNA damage that most severely affects the genomic integrity of the cell. It is essential that DNA DSBs are recognized and repaired efficiently, in particular prior to mitosis, to prevent genomic instability and eventually the development of cancer. To assess the pathways that are induced upon DNA DSBs, 14 human lymphoblastoid cell lines were challenged with bleomycin for 30 and 240 minutes to establish the fast and more prolonged response, respectively. The proteome of 14 lymphoblastoid cell lines was investigated to account for the variation amongst individuals. The primary DNA DSB response was expected to occur within the nucleus, therefore the nuclear extracts were considered. Differential analysis was performed using 2D DiGE and paired ANOVA statistics was used to recognize significant changes in time. Many proteins of which the levels in the nuclei changed statistically significant showed a fast response, i.e. within 30 minutes after bleomycin challenge. A significant number of these proteins could be assigned to known DNA DSB response processes, such as sensing DSBs (Ku70), DNA repair through effectors (HMG1), or cell cycle arrest at the G2/M phase checkpoint (14-3-3 zeta). Interestingly, the nuclear levels of all three proteins of the INHAT complex were reduced after 30 minutes of bleomycin challenge, suggesting that this complex might have a role in changing the chromatin structure, allowing DNA repair enzymes to gain access to DNA lesions.

INTRODUCTION

Decreased DNA stability is a key factor in the development of cancer. An impaired response to DNA damage, in particular to double strand breaks (DSBs), plays a role in the earliest stages of carcinogenesis (1). In addition, in inherited genetic syndromes such as *ataxia telangiectasia* and *Fanconi anemia*, genes that play a role in maintenance of DNA stability are affected and these syndromes are associated with a predisposition to cancer (2). Specific polymorphisms or low pathogenic mutations in these genes may subtly affect the function of the encoded enzymes, causing intermediate phenotypes associated with moderate increase in cancer risk (2-4). Nevertheless, not only genotype, but also exposure to environmental carcinogens is an important determinant for cancer risk. In particular, epithelial cells lining the body surfaces are continuously challenged by exogenous DNA-damaging agents. An example of what this can lead to are head and neck squamous cell carcinoma (HNSCC). These tumors arise in the mucosal linings of the upper aerodigestive tract in the 6^{th} and 7^{th} decade of life. The main

etiological factors are smoking and alcohol drinking that together have a synergistic effect. The risk for developing HNSCC is related to carcinogen exposure and for heavy smokers and drinkers the relative risk is as large as 20 (5). Notwithstanding, the large majority of smokers and alcohol drinkers do not develop HNSCC, indicating that individual cancer risk not solely depends on exposure but that this acts in concert with the capacity of individuals to deal with DNA damage induced by these exogenous agents (6-8).

Of the numerous DNA damage types occurring, cellular genome integrity is most severely affected by DSBs. Cells respond to DSBs by activating complex response pathways, including cell cycle arrest, DNA repair activation, and -in case of extensive damage- apoptosis (9, 10). There are two distinct and complementary mechanisms known for DNA DSB repair, i.e. homologous recombination and non-homologous end-joining (10). Much is already known about the signaling pathways that are induced upon the formation of DNA DSBs. The involvement of tumor suppressor genes such as p53, ATM, BRCA1, BRCA2, and of many other genes involved in the cell cycle checkpoint, DNA repair or apoptosis, has so far predominantly been investigated at the level of single proteins or separate signaling pathways. These signaling pathways have been studied at the level of mRNA to establish the effect of DSB induction, like the transcriptional response of lymphoblastoid cells to ionizing radiation (11-13). A more comprehensive analysis of the proteomic response to DNA DSB induction in cells, involving several human samples to enable appropriate statistics however, has never been reported.

Therefore, DNA damage was induced in human lymphoblastoid cells using bleomycin, which was selected since its mode of action resembles that of environmental carcinogens. Moreover, we are particularly interested in the response to DSBs, and bleomycin is more efficient in producing double strand breaks than ionizing radiation, (which is often used for this purpose) since it binds to and can cleave DNA at two opposing strands, resulting in a DSB (14). The ratio between double- and single strand breaks is 1:9 for bleomycin, and 1:100 for ionizing radiation (15). To identify both fast and more prolonged changes, the proteomic response to DNA DSB induction was determined at two time points. Samples were analyzed using two-dimensional difference in gel electrophoresis (2D DiGE), allowing the simultaneous analysis of many samples, which could not be easily performed using any of the alternative approaches in quantitative proteomics. The use of fluorescent dyes permitted analysis of small changes in protein expression both for relatively low and high protein levels, and the use of an internal standard to normalize the gel-to-gel variation. Proteins of which the nuclear levels changed statistically significant upon DNA DSB induction were identified using (tandem) mass spectrometric techniques. Our results showed that

many of the identified proteins are part of known DNA DSB response signaling pathways and that these proteins demonstrated a relatively fast response. Surprisingly, a fast decrease was detected in the relative levels of the three proteins forming the so-called 'inhibitor of acetyltransferases' or INHAT complex, of which involvement in DNA DSB response has not been described before.

EXPERIMENTAL PROCEDURES

Donor characteristics, cell culture and stimulation of cells

Blood was drawn from 14 individuals who took part in an epidemiological study evaluating risk factors for HNSCC. The study was approved by the Institutional Review Board of the VU University Medical Center, and written informed consent was obtained from all individuals. For each individual, a lymphoblastoid cell line was produced by immortalization with EBV, after which the phenotype was checked (7). These cell lines were cultured for a relatively short period of time (no more than 20 passages) in RPMI 1640 (with glutamine) medium-supplemented with 15% fetal bovine serum (BioWhittaker, Verviers, Belgium), 1% penicillin and streptomycin (Gibco-Invitrogen, Paisley, UK) and 0.1% 1M pyruvic acid (Sigma, Zwijndrecht, the Netherlands). Cells were incubated in a humidified incubator at 37°C in a 5% CO_2 in cultured flasks with filter caps (Nalgene-Nunc, Roskilde, Denmark).

Cells were challenged using bleomycin (Dagra Pharma, Diemen, the Netherlands) at a final concentration of 10 μM in H_2O to induce DSBs (16), and were harvested after 30 and 240 minutes, respectively. Control cells were harvested without any challenge.

Sample preparation and protein labeling

For harvesting, cells were pelleted and washed on ice using PBS containing 1 tablet/ 25 mL Complete Protease Inhibitor Cocktail (Roche Diagnostics, Almere, The Netherlands) and 1/100 (v/v) of Phosphatase Inhibitor Cocktails 1 and 2 (Sigma). Nuclei were purified using the protocol of Busch *et al.* (17). Subsequently, nuclear extracts were prepared basically as described by Valcarcel *et al.* (18). In short, purified nuclei were lysed in hypertonic buffer containing 20 mM HEPES pH 7.9, 420 mM KCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 20% glycerol, 1 mM DTT and protease and phosphatase inhibitors as mentioned above. After continuous stirring at 4°C for 1 hour, samples were centrifuged at 40,000 *g*. The supernatants, i.e. the nuclear extracts, were collected and stored at –80°C.

Nuclear extract protein concentrations were determined using the PlusOne 2D Quant Kit (Amersham Biosciences, Uppsala, Sweden). For each 2D

gel, 50 μg of nuclear extract proteins was precipitated using the PlusOne 2D Clean Up Kit (Amersham Biosciences) and was redissolved in 5-10 μL labeling buffer containing 30 mM Tris, pH 8.5, 7 M urea, 2 M thiourea and 4% CHAPS.

Nuclear protein samples were labeled as previously described (19). In brief, 50 μg of precipitated protein was labeled with 400 pmol of cyanine dye (Cy3 or Cy5, see Table 1 for the labeling scheme) during 30 minutes in the dark.

Table 1. *Setup of the 2D DiGE experiment to analyze 3x14 human nuclear extracts in a paired manner. Lymphoblastoid cells were stimulated with bleomycin for 0, 30, and 240 minutes, indicated with (t = 0), (t = 30') and (t = 240'), respectively. For each of the 28 gels, the Cy2 labeled internal standard consisted of an equal amount of all 42 samples that were analyzed. Other samples, of which the differential nuclear protein abundance had to be established, were labeled with Cy3 or Cy5 as indicated.*

2D gel #	Cy2 labeled sample	Cy3 labeled sample	Cy5 labeled sample
1	Internal standard	Nuclear extract (t = 0)	Nuclear extract (t = 30')
2	Internal standard	Nuclear extract (t = 0)	Nuclear extract (t = 240')

The internal standard, consisting of equal amounts of all 42 samples, was labeled with Cy2. The cyanine dyes contain an NHS-ester reactive group that covalently attaches to the epsilon amino group of protein lysine residues. Since the cyanine dyes are limiting in the reaction, on average only one lysine residue in labeled *per* protein. The labeling reaction was quenched by adding 1 μl of a 10 mM lysine solution. Both steps were performed at 4°C. Labeled samples were stored at -80°C for further analysis. CyDye labeling efficiency was verified by comparing the total density of a 1D SDS-PAGE image of a nuclear extract lysate with that of an *E. coli* lysate, prepared according to the description of the manufacturer.

Two-dimensional difference in-gel electrophoresis

After pooling the Cy2-labeled internal standard, Cy3- and Cy5-labeled samples according to Table 1, proteins were resolved according to their charge and size using two-dimensional gel electrophoresis. Conditions for 2D gel electrophoresis were as previously described (19). The cyanine dyes are designed such that dye-labeled lysines still contain a single positive charge and possess an added M_r of approximately 500 Da. This combination of size- and charge matching ensures an exact overlay of the protein patterns from the three differently labeled samples in a single 2D gel.

Image acquisition and statistical analysis

Because the three fluorescent Cy-dyes are spectrally distinct, each sample in a 2D gel image was independently scanned at the appropriate label-specific excitation and emission wavelengths using a Typhoon 9400 imager (Amersham Biosciences) equipped with narrow band pass filters. DeCyder software (version 5.01.01, Amersham Biosciences) was used for image analysis as described before (19). The number of estimated spots was set to 2,500. Detection and matching of protein spots required manual intervention to set landmarks on the gels to increase cross-gel matching accuracy. Data concerning the spot volume (with background subtracted) were collected for each spot, and then standardized abundance ratios were calculated for each pair of images (Cy3:Cy2, and Cy5:Cy2). These ratios were normalized to correct for differences in dye intensities, and subjected to statistical analysis, which was all performed within the DeCyder software. A paired analysis of variation (ANOVA) method was used, which assigned statistical significance to the differences in normalized protein abundance between the three different time points of bleomycin stimulation. Firstly, a change in protein spot intensity was only further regarded as relevant when it was detected in at least 80% of the 84 images. Secondly, the nuclear protein abundances were considered to be changed significantly when $p<0.05$. To account for genetic variation among individuals, 14 human lymphoblastoid cell lines were explored, which served as replicates. To establish the effect of considering less than 14 individuals, data sets from 2 up to 14 randomly chosen non-similar samples were evaluated with paired ANOVA statistics, using the same criteria as described for the evaluation of 14 samples, i.e. detection of the spot in more than 80% of the gels and $p<0.05$.

Staining, protein identification using mass spectrometry

As described before (19), gels were stained with silver after fluorescence detection, which enabled manual spot excision, and subsequently protein digestion was performed. MALDI-TOF MS peptide mass fingerprints were acquired on a Voyager DE-STR MALDI TOF mass spectrometer (Applied Biosystems, Nieuwerkerk a/d IJssel, the Netherlands) in the positive reflectron mode with delayed extraction, using the following settings: Accelerating voltage 20 kV, grid voltage 72%, guide wire 0.01%, delay time 150 nsec. Internal calibration was performed using trypsin auto-digest peaks. After baseline correction and noise filtering, spectra were de-isotoped and the resulting peak lists were searched against the SwissProt and NCBI databases for protein identity using the Mascot search engine. MS/MS spectra were acquired on a 4700 Proteomics Analyzer MALDI-TOF/TOF mass spectrometer (AB 4700 Proteomics Analyzer, Applied

Biosystems) This instrument is equipped with a 200 Hz Nd:YAG laser operating at 355 nm. Experiments were performed in a reflectron positive ion mode using delayed extraction. Typically, 2,000 shots per spectrum were acquired in the MS mode and 15,000 shots/spectrum in the MS/MS mode. Details of the peptide sequence coverage and Mascot score are given in the Supplementary Table.

Biochemical analysis of INHAT proteins

A core histone N-terminal consensus peptide, with sequence SGRGKAGKKGRKGAKTRQC, was immobilized onto Sulfolink beads (Pierce Biotechnology) at a concentration of 1 mg/mL. Affinity purifications were performed as described by Schneider *et al* (20). The immobilized peptide was fully acetylated during 1 hour at room temperature using NHS-acetyl in 200 mM Na_2CO_3, pH 8. After washing the beads, the experiment was repeated as described above. Eluted proteins were analyzed using 1D SDS-PAGE and visualized using silver staining. For Western blotting, nuclear proteins were resolved by SDS-PAGE and subsequently transferred to nitrocellulose (Bio-Rad, Veenendaal, the Netherlands), and probed with the primary anti-pp32A antibody (Abcam, Cambridge, UK) in block buffer (3% BSA, 0.1% Tween 20, v/v in PBS) and incubated overnight at 4^{o}C. A Cy5-linked secondary antibody (Abcam) was used to visualize the pp32A bands, fluorescence intensities were determined using the Typhoon 9400 fluorescence scanner (Amersham Biosciences), and subsequently normalized using the total protein intensity as detected with Coomassie Blue.

RESULTS

2D DiGE

The nuclear proteomic response to DNA double strand break induction was investigated in 14 human lymphoblastoid cell lines, using 2D DiGE. This approach permitted a simultaneous paired comparison of 56 samples, which were normalized against 28 internal standards (Table 1). Image analysis revealed that typically 2,100 protein spots were detected per gel, and that differences both in low- and high protein expression levels could be quantified over a linear dynamic range of 4-5 orders of magnitude. Moreover, the set of 84 gel images was highly reproducible, since over 60% of the spots were detected in all gels. Figure 1 shows a representative example of a post-silver stained 2D DiGE gel image, in which spot positions of proteins that will be discussed are indicated.

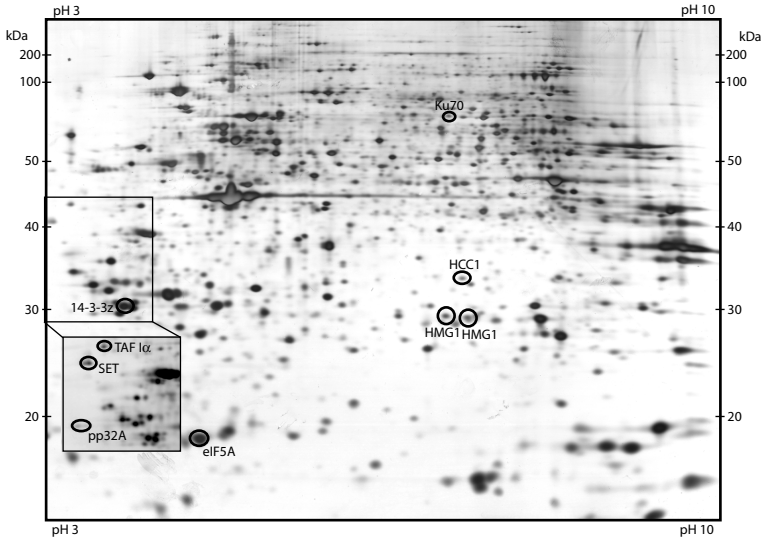


Figure 1. *Representative example of a post-silver stained 2D gel image from the analysis set. Note that every spot intensity appears as the sum of three individual sample spot intensities, since each gel contains three samples, labeled with separate Cy-Dyes. On average 2100 spots were separated on a non-linear pH 3-10 IPG strip in the first dimension and on a 12.5% SDS-PA gel in the second dimension. The selected subset of proteins, as listed in Table 2, is indicated in the image.*

Further characteristics of these indicated proteins are summarized in Table 2. The complete dataset of identified proteins can be found in the Supplementary Table. Although we restricted the paired ANOVA *p*-values to 0.05, the majority of *p*-values were much smaller than 0.05, i.e. for 50 proteins the *p*-values ranged from 9×10^{-18} to 0.01, for 16 proteins the *p*-values ranged from 0.01 to 0.025, and for 22 proteins the *p*-value ranged from 0.025 to 0.05. Hence, in total 88 nuclear extract proteins demonstrated a significant change in protein abundance, and remarkably, significant changes in protein abundance occurred rather fast, i.e. after 30 minutes of bleomycin challenge. Moreover, it was estimated how many lymphoblastoid cell lines would have been necessary for a study as described here, i.e. a study in which it is essential to account for genetic variation amongst individuals. As shown in Figure 2, it seems that for this data set of 2,100 detected spots, a more or less constant number of significantly changed protein levels is found when 9-14 cell lines are considered. Figure 2 also shows an evaluation of the consistency of the data sets, i.e. how many proteins are found that are in common with the final set found in 14 cell lines, 80% or

more proteins are in common with the final set for 9-14 cell lines, whereas 92% or more proteins are in common for 11-14 cell lines.

Table 2. *Selected subset of proteins that showed significantly changed nuclear levels upon DSB induction, and of which the relation with DNA DSBs is discussed in this paper. The regulation is either indicated by arrows (up and down to point out up-regulation or down-regulation, respectively) or by an equal sign (for 'no change'). The complete list of differential proteins that were identified, and other relevant data are summarized in the Supplementary Table. All proteins were identified using MALDI-TOF peptide mass fingerprinting and MALDI-TOF/TOF peptide sequencing.*

Protein	Regulation*		Accession number†	Mass‡ (Da)	pI**	SC†† (%)
	A	B				
Sensors of DNA damage						
ATP-dependent DNA helicase II, 70 kDa (Ku70)	▲	▲	P12956	70084	6.23	31
Effectors						
Inhibitor of acetyltransferases complex						
Template activating factor 1α (TAF 1α)	▼	=	Q01105-1	33469	4.23	24
SET protein (TAF 1β)	▼	=	Q01105-2	32084	4.12	21
Acidic nuclear phosphoprotein pp32 (pp32A)	▼	=	P39687	28585	4.50	41
DNA-interacting proteins						
High mobility group protein 1 (HMG1)	▼	=	P09429	25049	5.62	41
High mobility group protein 1 (HMG1)	▼	=	P09429	25049	5.62	54
Nuclear protein Hcc-1 (HCC1)	▲	▲	P82979	23713	6.10	44
Cell cycle						
Eukaryotic translation initiation factor 5A (eIF5A)	▼	=	P63241	17049	5.08	61
14-3-3 zeta/delta (14-3-3)	▼	▲	P63104	27899	4.73	28

* Regulation of protein expression levels; A: regulation between 0 and 30 minutes of bleomycin stimulation, B: regulation between 30 and 240 minutes of bleomycin stimulation.

† Accession number according to entries in the SwissProt Database,

‡ Theoretical values, mass in Dalton

** Theoretical values

†† Sequence coverage, obtained by a combination of peptide mass fingerprinting and ToF/ToF peptide sequencing

Supplementary Table. *Characteristics of all proteins that showed a significant change in nuclear protein levels in time in response to DNA damage induction. For each protein the following characteristics are given; the protein name and accession numbers as given in the Swiss-Prot and the NCBI non-redundant databases, the theoretical molecular weight and pI, the average ratios in protein level showing the changes after 30 and after 240 minutes of bleomycin challenge, the p-value from the paired 1-ANOVA, the number of peptides used for identification of the protein, the Mascot score and the percentage sequence coverage.* **This table can be found on the supplementary CD-rom**

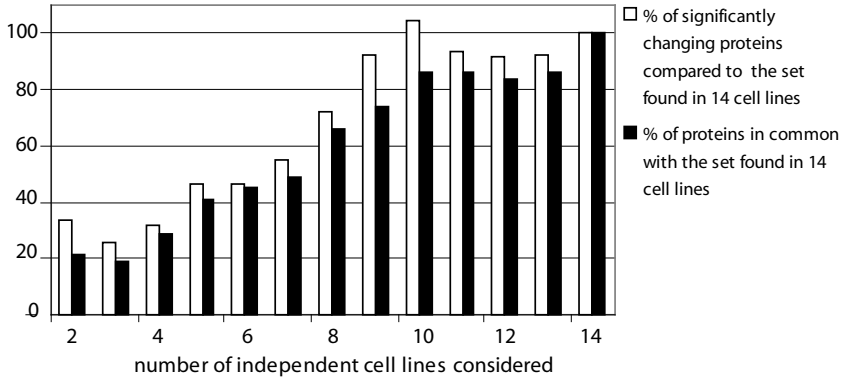


Figure 2. *Assessment of the effect of the number of lymphoblastoid cell lines taken into consideration on the number of protein levels that are significantly changed, and the consistency of the data sets. The 14 cell lines used for this study were taken as reference. For this analysis, 2-14 independent randomly chosen cell lines were subjected to paired 1-way ANOVA statistics. The white bars represent the percentage of proteins that were significantly changing compared to the reference, which is relatively constant when 9-14 cell lines are considered. The black bars indicate the consistency of the data sets, i.e. the number of proteins in common with the reference, showing that 80% or more proteins are in common with the final set for 9-14 cell lines, whereas 92% or more proteins are in common for 11-14 cell lines.*

Protein identification

Using MALDI peptide mass fingerprinting and MALDI-TOF/TOF sequence analysis, 56 significantly changed proteins could be identified, of which the results are summarized in Table 2 and the Supplementary Table. The low expression level of the 32 other significantly changed proteins hampered their identification, even when analyzing spots from preparative gels. This shows that the sensitivity of fluorescent detection exceeds that of MS detection. An example of the protein identification of HMG1 using mass spectrometry is depicted in Figure 3, showing a typical flow scheme for the identification of the high mobility group protein 1 (HMG1; the spot of the most basic isoform in Figure 1). The paired ANOVA statistical analysis showed that for this spot the nuclear protein level significantly changed over three time points in all samples. The protein was identified through its peptide mass fingerprint, and identification was confirmed using the sequence information in the MS/MS spectra, of which an example is shown in Figure 3B and 3C, respectively.

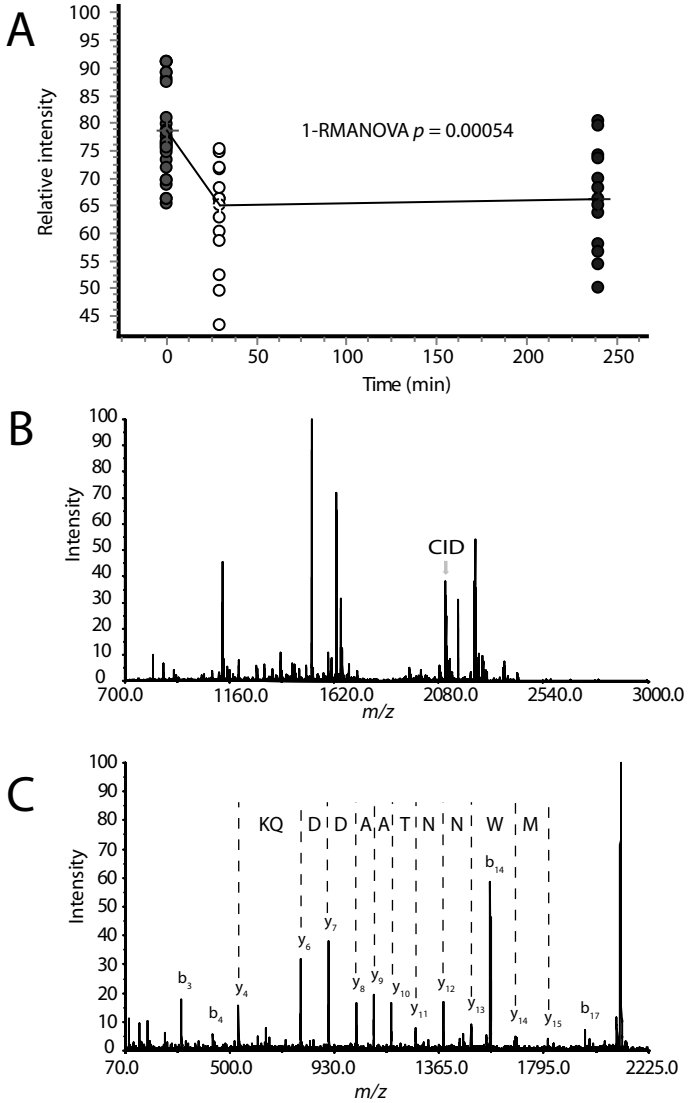


Figure 3. *Flow scheme for the mass spectrometric identification of a differentially regulated protein, high mobility group 1 or HMG1 protein, from a 2D gel. A) Statistical analysis, i.e. paired 1-way ANOVA, showed that the protein significantly changed in abundance over three time points in all samples. The corresponding protein spot was excised from the gel and subsequently digested using trypsin. B) The peptide mass fingerprint was used to identify the protein. C) To confirm the identification, a peptide was selected for collision-induced dissociation (CID), from which sequence information was retrieved.*

Response to DNA DSB induction

All proteins of which the expression level in the nuclei significantly changed upon bleomycin challenge were classified according to their function, which was based on the information present in the SwissProt [http://us.expasy.org/sprot/] and the NCBI non redundant [http://www.ncbi.nlm.nih.gov/] databases. Table 2 shows the proteins classified as involved in DNA repair (ATP-dependent DNA helicase II or Ku70), in DNA interaction (template activating factor Iα, SET protein, acidic nuclear phosphoprotein pp32, high mobility group protein 1, and nuclear protein Hcc-1), and involved in cell cycle regulation (14-3-3 zeta/delta, eukaryotic translation initiation factor 5A). Significant changes were also found for protein levels known to be involved in several other processes such as in the oxidative stress response, these were however not directly related to DNA DSBs (see the Supplementary Table). High mobility group protein 1, or HMG1, is an example of a protein that was identified in more than one distinct protein spot, differing in pI, which may implicate that post-translational modification occurred. Unfortunately, the mass spectra of these samples did not reveal the type of post-translational modification.

Response of INHAT complex to DNA DSBs

A remarkable response to the induction of DNA DSBs was the decrease in protein levels of the template activating factor Iα (TAF 1α), the SET protein (TAF 1¤), and the acidic leucine-rich nuclear phosphoprotein pp32 A (pp32A), proteins that form the 'inhibitor of acetyltransferases' (INHAT) complex (21, 22). The decrease in nuclear levels of this complex suggests that it might be involved in the response to DNA DSBs through regulation of chromatin structure. Even though this complex is known to be involved in HAT-dependent transcriptional regulation (23, 24), a role in DNA DSB damage response has not been described before. It has been found that histone N-terminal tails are central to regulation of chromatin structure, as they form a binding platform for multiple protein complexes such as the INHAT complex. Furthermore, binding to the N-terminal tail was reported to be affected by various modifications within the histone N-terminal tail sequence (20). Therefore, our results concerning the INHAT complex proteins were further investigated using the same lymphoblastoid nuclear extracts. To establish whether the INHAT complex binds preferably to non-acetylated histones, an immobilized peptide with the sequence SGRGKAGKKGRKGAKTRQC was used as bait, which is a sequence considered as histone N-terminal 'consensus' peptide. Both the non-acetylated and the acetylated form of this histone N-terminal 'consensus' peptide were used in

separate experiments. Specifically bound proteins were separated and identified using 1D SDS-PAGE and mass spectrometry. As can be concluded from Figure 4A, the INHAT complex proteins were detected in the eluate and not in the flow through for the non-acetylated histone N-terminal 'consensus' peptide.

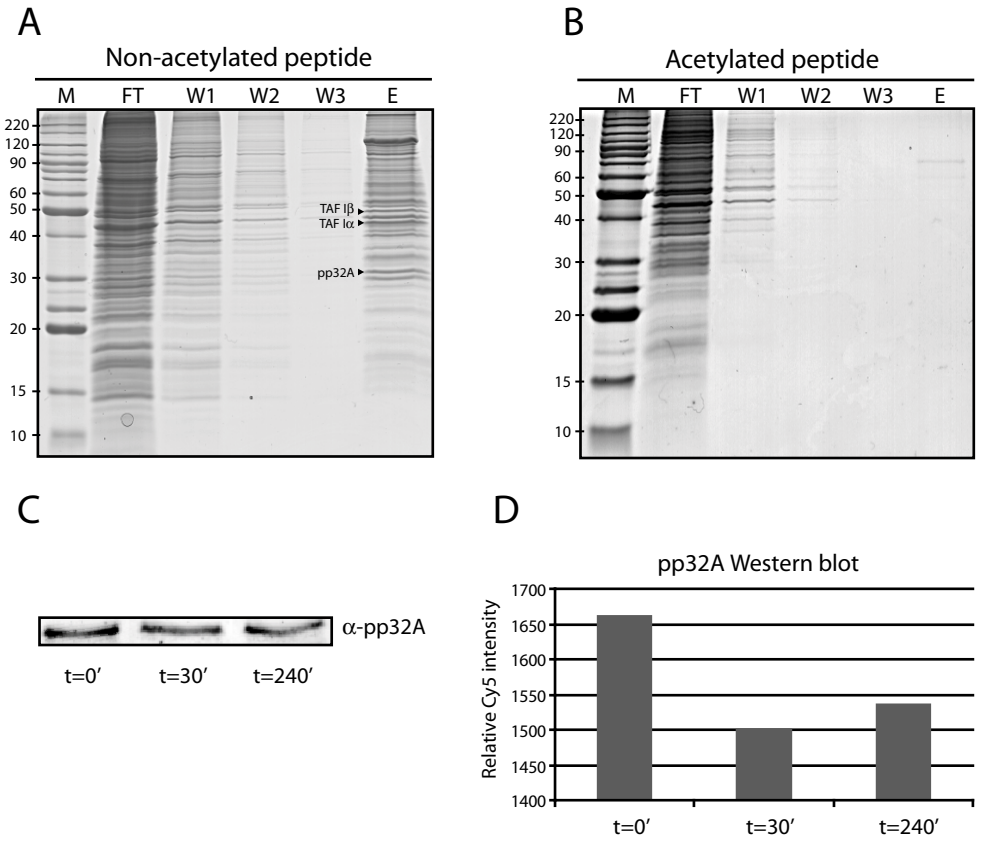


Figure 4. *Biochemical analysis of specific binding of INHAT complex proteins to a histone N-terminal 'consensus' peptide. A) A synthetic peptide reflecting a histone N-terminal 'consensus' sequence SGRGKAGKKGRKGAKTRQC was immobilized through the C-terminal cysteine onto Sulfolink beads. Proteins that specifically bound to this peptide were separated using 1D-SDS-PAGE (M, molecular weight marker; FT, flow-through; W1/2/3, wash fractions 1, 2 and 3; E, eluate fraction). Mass spectrometric analysis of this fraction revealed binding of the three proteins of the INHAT complex (arrows). B) After complete acetylation of the immobilized peptide, the experiment was repeated. This showed that binding of INHAT to the non-acetylated histone N-terminal tail peptide was specific because these proteins were detected in the flow-through and not in the eluate fraction. C) Western blot results of the regulation of pp32A, one of the INHAT proteins. D) Relative quantitation of pp32A protein levels.*

Binding to this peptide was indeed specific, since the INHAT complex proteins did not bind to the acetylated histone N-terminal 'consensus' peptide. This is shown in Figure 4B, where the INHAT complex proteins were detected in the flow through and not in the eluate fraction. This is in agreement with results obtained by Schneider *et al* (20) who found reduced binding of INHAT proteins to histone N-terminal tails upon (partial) modification of amino acids in the N-terminus. The protein band that appeared in the eluate fraction was identified as the NAD-dependent deacetylase sirtuin-1, which is known to deacetylate (and bind) acetylated histones *in vitro*. This protein was not identified in the eluate from the non-acetylated immobilized peptide. In addition, the 2D DiGE quantification of the INHAT complex proteins in the nuclear extracts was validated for pp32A using Western blotting. The example depicted in Figure 4C, showing the densitograms, and 4D, showing the relative quantities, confirmed that the nuclear levels of pp32A decreased after 30 and 240 minutes of bleomycin challenge compared to the reference (t=0).

DISCUSSION

DNA DSBs are highly cytotoxic and must be recognized and repaired efficiently to prevent genomic instability and eventually the development of cancer. The signaling pathways that are activated upon DNA DSB induction were investigated by challenging 14 human lymphoblastoid cell lines with bleomycin, and subsequent monitoring of the proteomic response in time. In the relatively small time span of 30–240 minutes of bleomycin challenge it is expected that transcriptional or translational control plays a minor role, since in particular in higher eukaryotes correct splicing and folding are time consuming processes. Instead, signaling at these time-intervals will mainly be controlled by post-translational modifications, such as protein phosphorylation. This can trigger the formation of protein-protein interactions or shuttling of proteins from the nucleus to the cytoplasm. The latter mechanism is described for example for regulation of BRCA1, which has been found to be regulated through nuclear export in response to DNA damage induced by ionizing irradiation (25).

Response of proteins known to be involved in DNA DSBs

A considerable fraction of the differential nuclear proteins was found to be involved in DNA DSB response pathways. The first step in the DNA damage response is the recognition of lesions in the DNA by specific sensor proteins (9, 10). The DNA repair machinery started relatively quickly, because the ATP-dependent DNA helicase II or Ku70 was one of the proteins of which the nuclear

level increased within 30 minutes of DSB induction. This important sensor and initiator of DNA repair is part of a heterodimer consisting of Ku70 and Ku80, of which Ku70 is known to bind DNA ends in a non-sequence specific manner and starts DNA repair through non-homologous end joining (26). In the subsequent step, sensor proteins recruit signal transducers such as the *ataxia telangiectasia* mutated (ATM) kinase (27), its homologue *ataxia telangiectasia* and Rad3 related protein (ATR, (28), the DNA-dependent protein kinase (DNA-PK) (29), p53 (30), BRCA1 (31), Chk1 and Chk2 (32). These signal transducer proteins were either not found to be differentially present in the nuclei or they were not detected, the latter since these regulatory proteins have a high molecular weight (ATM/ATR) or are very low abundant and unstable (p53) and thus hard to visualize on a 2D gel.

In turn, the signal transducers activate a broad range of effector proteins, for example those involved in cell cycle arrest (33, 34). A relatively fast response was found for the effector protein 14-3-3zeta/delta, of which the zeta isoform is known to be involved in cell cycle regulation through binding to Cdc25 (35, 36). Cdc25 is a phosphatase that dephosphorylates Cdc2, thereby positively regulating entry into mitosis. However, upon induction of DNA damage, Cdc25 is phosphorylated to create a 14-3-3 zeta binding site. After binding, 14-3-3-bound Cdc25 is sequestered from the nucleus into the cytosol where it cannot act on Cdc2, thus mitosis is prevented, and as a result the cell cycle is halted at the G_2/M checkpoint (35, 36). Additionaly, a reduction in level of eIF5A1 was detected after 30 minutes, a protein of which the function is related to the checkpoint at the G_1/S transition. This may indicate an arrest of the cell cycle prior to the start of the S phase, to prevent that damaged DNA will be replicated (37, 38). Recently, it has been found that a complex of eIF5A and syntenin has a function in regulation of p53 and p53-dependent apoptosis (39).

Another category of effector proteins consists of those involved in DNA repair. High mobility group 1 protein (HMG1) was found to be reduced in abundance in the nuclear extract after bleomycin stimulation. HMG1 has been described to enhance intra- and intermolecular ligations of cohesive-ended and blunt-ended DNA through its DNA binding properties, and it was suggested that HMG1 is involved in the final ligation step in DNA end-joining processes (40, 41). It has also been found that HMG1 is a direct activator of the DNA binding capacities of p53. This interaction has been found to be independent of the C-terminal tail of p53, a region known to interact with other activators (42). Another DNA binding protein is the nuclear protein Hcc-1 that showed an increase in abundance only after 240 minutes of bleomycin challenge. Not much is known about the exact function of this protein, although a PSI-BLAST search showed that the first 42 amino acids of the protein form a SAP domain (43). This

type of domain is a putative DNA binding motif that is known to be involved in transcription regulation and DNA repair (44).

Response of INHAT complex to bleomycin challenge

Strikingly, it was established that the nuclear levels of the tripartite INHAT complex proteins were reduced after 30 minutes of bleomycin challenge. It has been suggested that this complex inhibits p300/CBP and PCAF-mediated acetylation of histones, most likely by masking the accessibility of the histone lysines (22-24). When levels of the INHAT complex proteins are reduced, the basic N-terminal ends of the core histones become accessible to p300/CBP. Histone acetylation by p300/CBP is an important mechanism in chromatin dynamics during transcription, since this results in a decreased density of packed DNA, which allows transcription factors to enter DNA and genes to be transcribed (45). The fact that the nuclear levels of all these three proteins of the INHAT complex were reduced after 30 minutes of DNA damage induction may indicate that this complex is not only involved in transcriptional regulation, but plays a similar role in response to DNA DSBs, as illustrated in Figure 5.

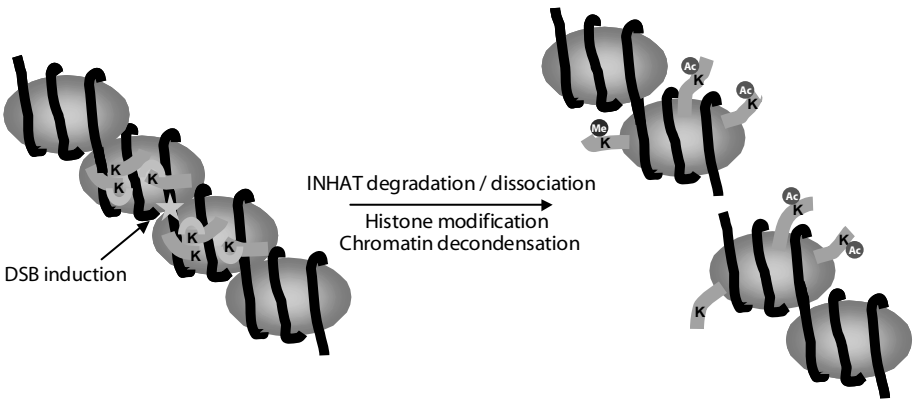


Figure 5. *Proposed mechanism of chromatin remodeling in the DNA damage response. Upon induction of a DSB (indicated by a star), the INHAT complex (indicated by the U-shaped structures) dissociates from the histone N-termini that are exposed from of the nucleosome core. This allows modification of the N-termini, such as methylation and acetylation (indicated by the closed circles) by histone methyl- and acetyltransferases, which in turn leads to chromatin decondensation, opens up chromatin and increases accessibility for DNA repair factors, like Ku70 and DNA-PK.*

Comparison to other proteomics studies

Even though the effect of DNA damage has been studied extensively, the response to DNA DSB inducing agents as efficient as bleomycin was never

investigated using a proteomics approach. In contrast, the proteomic responses to ultraviolet (UV) irradiation of HeLa cells (46) and to ionizing irradiation induction of L929 cells (47) have already been investigated. Some of the proteins that were found to be differentially regulated upon UV irradiation over a time course of 24 hours were also found in our study, like eIF5A, lactate dehydrogenase, and proteasome subunits, but the main effect was detected to occur on DNA replication at the G_1/S boundary (46). The L929 cells were challenged with 6 Gy ionizing radiation, and after 72 hours differential expression of 47 proteins was detected. The identified proteins participate either in protective and reparative cell responses or induction of apoptosis. Like in our study, eIF5A and several subunits of the proteasome, 14-3-3 zeta, transketolase, and protein disulfide isomerase were found to be differentially regulated (47).

In conclusion, the use of 2D DiGE allowed investigation of the proteomic response of 14 independent human biological replicates to DNA DSBs induced by bleomycin. It was established that for this study of human samples at least 9 independent replicates are needed tot find a rather constant and consistent number of significant changes in nuclear protein levels. Many of the proteins that changed in nuclear level are known to be involved in repair of DNA DSBs. Although it is known which pathways are involved to a large extent, it is still not known how DNA repair factors can gain access to DNA lesions within chromatin, prior to repair of damaged DNA. The significance of chromatin remodeling in DNA repair has been reported in several studies (48-50). Intriguingly, protein levels of the tripartite INHAT complex were found to be significantly reduced upon DNA DSB induction, suggesting that these inhibitors of chromatin modifiers not only play a role in transcription, but also in DNA repair.

ACKNOWLEDGMENTS

We are grateful to Mirjam Damen and Judith Barends for their technical assistance in protein identification through mass spectrometry. We want to thank Dr. Dirk Rijkers of the Medicinal Chemistry group, Utrecht University, for providing us with the histone N-terminal 'consensus' peptide. This project was supported by the Netherlands Proteomics Centre.

REFERENCES

1. Gorgoulis, VG, Vassiliou, LV, Karakaidos, P. (2005) Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* **434**, 907-13.
2. Vogelstein, B and Kinzler, KW. (2004) Cancer genes and the pathways they control. *Nat Med* **10**, 789-99.
3. Joenje, H and Patel, KJ. (2001) The emerging genetic and molecular basis of Fanconi anaemia. *Nat Rev Genet* **2**, 446-57.
4. Benhamou, S, Tuimala, J, Bouchardy, C, et al. (2004) DNA repair gene XRCC2 and XRCC3 polymorphisms and susceptibility to cancers of the upper aerodigestive tract. *Int J Cancer* **112**, 901-4.
5. Brugere, J, Guenel, P, Leclerc, A, and Rodriguez, J. (1986) Differential effects of tobacco and alcohol in cancer of the larynx, pharynx, and mouth. *Cancer* **57**, 391-5.
6. Cloos, J, Nieuwenhuis, EJ, Boomsma, DI, et al. (1999) Inherited susceptibility to bleomycin-induced chromatid breaks in cultured peripheral blood lymphocytes. *J Natl Cancer Inst* **91**, 1125-30.
7. Cloos, J, Temmink, O, Ceelen, M, et al. (2002) Involvement of cell cycle control in bleomycin-induced mutagen sensitivity. *Environ Mol Mutagen* **40**, 79-84.
8. Cloos, J, Leemans, CR, van der Sterre, ML, et al. (2000) Mutagen sensitivity as a biomarker for second primary tumors after head and neck squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev* **9**, 713-7.
9. Jackson, SP. (2002) Sensing and repairing DNA double-strand breaks. *Carcinogenesis* **23**, 687-96.
10. Khanna, KK and Jackson, SP. (2001) DNA double-strand breaks: signaling, repair and the cancer connection. *Nat Genet* **27**, 247-54.
11. Tusher, VG, Tibshirani, R, and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **98**, 5116-21.
12. Jen, KY and Cheung, VG. (2003) Transcriptional response of lymphoblastoid cells to ionizing radiation. *Genome Res* **13**, 2092-100.
13. Rieger, KE and Chu, G. (2004) Portrait of transcriptional responses to ultraviolet and ionizing radiation in human cells. *Nucleic Acids Res* **32**, 4786-803.
14. Abraham, AT, Zhou, X, and Hecht, SM. (2001) Metallobleomycin-mediated cleavage of DNA not involving a threading-intercalation mechanism. *J Am Chem Soc* **123**, 5167-75.
15. Cloos, J, Gille, JJ, Steen, I, et al. (1996) Influence of the antioxidant N-acetylcysteine and its metabolites on damage induced by bleomycin in PM2 bacteriophage DNA. *Carcinogenesis* **17**, 327-31.
16. Adema, AD, Cloos, J, Verheijen, RH, Braakhuis, BJ, and Bryant, PE. (2003) Comparison of bleomycin and radiation in the G2 assay of chromatid breaks. *Int J Radiat Biol* **79**, 655-61.
17. Busch, H and Daskal, Y. (1977) Methods for isolation of nuclei and nucleoli. *Methods Cell Biol* **16**, 1-43.
18. Valcarcel, R and Stunnenberg, HG. (1996) Retinoid-dependent in vitro transcription. *Methods Enzymol* **274**, 149-61.
19. Kolkman, A, Dirksen, EH, Slijper, M, and Heck, AJ. (2005) Double standards in quantitative proteomics: direct comparative assessment of difference in gel electrophoresis and metabolic stable isotope labeling. *Mol Cell Proteomics* **4**, 255-66.
20. Schneider, R, Bannister, AJ, Weise, C, and Kouzarides, T. (2004) Direct Binding of INHAT to H3 Tails Disrupted by Modifications. *J Biol Chem* **279**, 23859-62.
21. Seo, SB, Macfarlan, T, McNamara, P, et al. (2002) Regulation of histone acetylation and transcription by nuclear protein pp32, a subunit of the INHAT complex. *J Biol Chem* **277**, 14005-10.
22. Seo, SB, McNamara, P, Heo, S, et al. (2001) Regulation of histone acetylation and transcription by INHAT, a human cellular complex containing the set oncoprotein. *Cell* **104**, 119-30.

23. Gamble, MJ, Erdjument-Bromage, H, Tempst, P, Freedman, LP, and Fisher, RP. (2005)The histone chaperone TAF-I/SET/INHAT is required for transcription in vitro of chromatin templates. *Mol Cell Biol* **25**, 797-807.
24. Kutney, SN, Hong, R, Macfarlan, T, and Chakravarti, D. (2004) A signaling role of histone binding proteins and INHAT subunits pp32 and Set/TAF-Ibeta in integrating chromatin hypoacetylation and transcriptional repression. *J Biol Chem* **279**, 30850-30855
25. Feng, Z, Kachnic, L, Zhang, J, Powell, SN, and Xia, F. (2004) DNA damage induces p53-dependent BRCA1 nuclear export. *J Biol Chem* **279**, 28574-84.
26. Hopfner, KP, Putnam, CD, and Tainer, JA. (2002) DNA double-strand break repair from head to tail. *Curr Opin Struct Biol* **12**, 115-22.
27. Andegeko, Y, Moyal, L, Mittelman, L, et al. (2001) Nuclear retention of ATM at sites of DNA double strand breaks. *J Biol Chem* **276**, 38224-30.
28. Abraham, RT. (2001) Cell cycle checkpoint signaling through the ATM and ATR kinases. *Genes Dev* **15**, 2177-96.
29. Smith, GC and Jackson, SP. (1999) The DNA-dependent protein kinase. *Genes Dev* **13**, 916-34.
30. Liu, Y and Kulesz-Martin, M. (2001) p53 protein at the hub of cellular DNA damage response pathways through sequence-specific and non-sequence-specific DNA binding. *Carcinogenesis* **22**, 851-60.
31. Venkitaraman, AR. (2001) Functions of BRCA1 and BRCA2 in the biological response to DNA damage. *J Cell Sci* **114**, 3591-8.
32. Iliakis, G, Wang, Y, Guan, J, and Wang, H. (2003) DNA damage checkpoint control in cells exposed to ionizing radiation. *Oncogene* **22**, 5834-47.
33. Cloos, J, Reid, CB, van der Sterre, ML, et al. (1999) A comparison of bleomycin-induced damage in lymphocytes and primary oral fibroblasts and keratinocytes in 30 subjects. *Mutagenesis* **14**, 87-93.
34. Holgersson, A, Heiden, T, Castro, J, et al. (2005) Different G2/M accumulation in M059J and M059K cells after exposure to DNA double-strand break-inducing agents. *Int J Radiat Oncol Biol Phys* **61**, 915-21.
35. Qi, W and Martinez, JD. (2003) Reduction of 14-3-3 proteins correlates with increased sensitivity to killing of human lung cancer cells by ionizing radiation. *Radiat Res* **160**, 217-23.
36. Chan, TA, Hermeking, H, Lengauer, C, Kinzler, KW, and Vogelstein, B. (1999) 14-3-3Sigma is required to prevent mitotic catastrophe after DNA damage. *Nature* **401**, 616-20.
37. Hanauske-Abel, HM, Slowinska, B, Zagulska, S, et al. (1995) Detection of a sub-set of polysomal mRNAs associated with modulation of hypusine formation at the G1-S boundary. Proposal of a role for eIF-5A in onset of DNA replication. *FEBS Lett* **366**, 92-8.
38. Park, MH, Lee, YB, and Joe, YA. (1997) Hypusine is essential for eukaryotic cell proliferation. *Biol Signals* **6**, 115-23.
39. Li, AL, Li, HY, Jin, BF, et al. (2004) A novel eIF5A complex functions as a regulator of p53 and p53-dependent apoptosis. *J Biol Chem* **279**, 49251-8.
40. Nagaki, S, Yamamoto, M, Yumoto, Y, et al. (1998) Non-histone chromosomal proteins HMG1 and 2 enhance ligation reaction of DNA double-strand breaks. *Biochem Biophys Res Commun* **246**, 137-41.
41. Yamanaka, S, Katayama, E, Yoshioka, K, et al. (2002) Nucleosome linker proteins HMGB1 and histone H1 differentially enhance DNA ligation reactions. *Biochem Biophys Res Commun* **292**, 268-73.
42. Jayaraman, L, Moorthy, NC, Murthy, KG, et al. (1998) High mobility group protein-1 (HMG-1) is a unique activator of p53. *Genes Dev* **12**, 462-72.
43. Aravind, L and Koonin, EV. (2000) SAP - a putative DNA-binding motif involved in chromosomal organization. *Trends Biochem Sci* **25**, 112-4.
44. Hashii, Y, Kim, JY, Sawada, A, et al. (2004) A novel partner gene CIP29 containing a SAP domain with MLL identified in infantile myelomonocytic leukemia. *Leukemia* **18**, 1546-8.

45. van Holde, K and Zlatanova, J. (1996) What determines the folding of the chromatin fiber? *Proc Natl Acad Sci U S A* **93**, 10548-55.
46. Decker, ED, Zhang, Y, Cocklin, RR, Witzmann, FA, and Wang, M. (2003) Proteomic analysis of differential protein expression induced by ultraviolet light radiation in HeLa cells. *Proteomics* **3**, 2019-27.
47. Szkanderova, S, Hernychova, L, Kasalova, I, et al. (2003) Proteomic analysis of radiation-induced alterations in L929 cells. *Folia Biol (Praha)* **49**, 15-25.
48. Verger, A and Crossley, M. (2004) Chromatin modifiers in transcription and DNA repair. *Cell Mol Life Sci* **61**, 2154-62.
49. Wuebbles, RD and Jones, PL. (2004) DNA repair in a chromatin environment. *Cell Mol Life Sci* **61**, 2148-53.
50. Peterson, CL and Cote, J. (2004) Cellular machineries for chromosomal DNA repair. *Genes Dev* **18**, 602-16.

Chapter 4

Assessing genetic susceptibility to head and neck squamous cell carcinoma from a proteomics perspective

Eef H.C. Dirksen[1], Jacqueline Cloos[2], Boudewijn J.M. Braakhuis[2], Ruud H. Brakenhoff[2], Albert J.R. Heck[1], Monique Slijper[1]

[1] Department of Biomolecular Mass Spectrometry, Utrecht Institute for Pharmaceutical Sciences, Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands

[2] Section Tumor Biology, Department of Otolaryngology/Head-Neck Surgery, VU University Medical Center, Amsterdam, The Netherlands.

ABSTRACT

Tobacco smoking and alcohol abuse are important risk factors for the development of tumors in the mucosal linings of the upper aerodigestive tract, such as head and neck squamous cell carcinomas (HNSCCs). Even though cancer risk is related to exposure, the large majority of smokers and alcohol drinkers do not develop HNSCC. It appears that individual cancer risk is determined by the interplay between exposure to these exogenous agents and the capacity of subjects to handle DNA damage. Hypersensitivity to chromatid breaks after exposure to bleomycin or γ-irradiation is regarded as biomarker of susceptibility to multiple cancer types, including HNSCC, and a decreased G_2/M block is associated with the hypersensitive phenotype. The exact pathways and involved proteins that could explain this phenotype however, are not known. Therefore, the response to bleomycin exposure (for 30 and 240 minutes) was investigated in the nuclear proteome of lymphoblastoid cell lines of which seven exhibit a normally sensitive and seven a hypersensitive phenotype. A total of seventeen nuclear protein levels differed significantly (p<0.01; two-sided) either between phenotypes or as interaction between both phenotype and time of bleomycin exposure. Several of these proteins take part in one or more complexes involved in known pathways of the DNA damage response. In particular, the aberrant regulation of HMG2, which is found at major intersections of these pathways, may form an important factor in enhanced susceptibility to DNA damage.

BRIEF COMMUNICATION

Maintenance of DNA integrity is essential for accurate transcription and correct transmission of genetic information. Consequently, decreased DNA stability is associated with development of cancer. This is most prominent with inherited disorders such as in *ataxia telangiectasia* (AT) and other hereditary cancer syndromes, often characterized by genomic instability and an increased risk for cancer development (1). Most sporadic cancers arising in the epithelial linings of the body however, develop as a result of the interplay between the exposure to exogenous carcinogenic agents and inherited susceptibility. This is well-illustrated by squamous cell carcinomas in the head and neck that arise as result of smoking and alcohol abuse, but only in a subgroup of exposed individuals, indicating that some individuals are able to handle high carcinogenic exposure while remaining free of cancer.

Susceptibility has a hereditary basis and can be measured in various cell types using model compounds such as bleomycin, which is a clastogenic

compound that causes double strand breaks in DNA. Blood cells (or lymphoblastoid cell lines thereof) from head and neck squamous cell carcinoma (HNSCC) patients, and particularly from those who developed multiple tumors, are more sensitive to the effects of DNA double strand breaks caused by bleomycin than the cells from control subjects (2,3). Accordingly, sensitivity to bleomycin-induced DNA damage can be considered a biomarker of cancer susceptibility. A subsequent study with monozygotic and dizygotic twins confirmed that the susceptibility for bleomycin induced DNA damage has a strong heritable basis (4). Functional studies revealed a relation with a decreased G_2/M block, causing so called damage-resistant proliferation (5). However, a clear molecular basis of this hypersensitive phenotype has not been established. Therefore, a comprehensive proteome analysis was performed to obtain insight in the underlying pathways.

Lymphoblastoid cell lines of seven hypersensitive and seven normal control subjects were cultured and subsequently exposed to bleomycin for 0 (controls), 30 or 240 minutes. Cell lines were as used previously (6,7). Extraction, separation of nuclear proteins using two dimensional difference in gel electrophoresis, and protein identifications were performed as described before (7). Approximately 2100 protein spots were detected in each of the 56 2D gels; their relative protein levels were determined using DeCyder software. A two-way analysis of variance (2-ANOVA) was conducted considering phenotype (condition 1) and time of bleomycin challenge (condition 2). Here, only significant differences in condition 1 and significant interaction between condition 1 and 2 were taken into account (p<0.01, two-sided). Using these stringent criteria, the nuclear protein levels of 17 proteins differed significantly (Table 1). These could be subdivided into 11 decreased and 4 increased in hypersensitive subjects compared to normal control subjects. Two proteins exhibited significant interaction values in protein levels. Some of the differential proteins are known to be involved in the immediate response to DNA damage, involving cell cycle regulation, cell proliferation and DNA repair. Moreover, the general response to bleomycin induction was studied previously (7), and both data sets share significantly changing proteins (indicated with '*' in Table 1). Five of these play an important role in DNA damage response pathways, suggesting that aberrant regulation of these proteins could play an essential role in the hypersensitive phenotype.

Table 1. *Proteins found to be differentially regulated between the normal control (N) and the hypersensitive (H) phenotype using a 2D gel/mass spectrometry based proteomics approach. The accession numbers correspond to those in the SwissProt database (http://www.expasy.org/sprot/).*

Protein name	**Accession number**	**2-ANOVA-S**	**2-ANOVA-I interaction**
Nuclear protein levels in normal control higher than in hypersensitive phenotype	(SwissProt)	*p*-value	*p*-value
ATP-dependent DNA helicase II, 80 kDa subunit	P13010	0.0037	
Heat shock cognate 71 kDa protein	P11142	0.0085	
Adenylyl cyclase-associated protein 1*	Q01518	0.0047	
Septin 2 (NEDD5 homologue)*	Q15019	0.006	
hnRNP A2/B1	P22626	7.80E-05	
PNAS-139	Q9BXV5	2.30E-05	
GTP-binding nuclear protein Ran	P62826	0.0016	
Rho GDP-dissociation inhibitor 2	P52566	0.0027	
Peroxiredoxin 2	P32119	0.0042	
Nucleoside diphosphate kinase A*	P15531	0.01	
Ubiquitin-conjugating enzyme E2 N	P61088	0.0052	
Nuclear protein levels in hypersensitive phenotype higher than in normal control			
Eukaryotic initiation factor 4A-I*	P60842	0.0096	
Thioredoxin-like protein 2	O76003	0.0063	
Acidic leucine-rich nuclear phosphoprotein 32A*	P39687	0.00037	
Rap1B, member of the Ras oncogene family	Q6DCA1	0.00032	
Interaction: regulation in time, and between normal control *versus* hypersensitive phenotype			
Heterogeneous nuclear ribonucleoprotein C1/C2	P07910		0.0084
High mobility group protein 2	P26583		0.006

Abbreviations: 2-ANOVA-S: *p*-value for the statistical significance between the phenotypes; 2-ANOVA-I: *p*-value for the statistical significance of the interaction between time and phenotype. *Protein was also significantly regulated in response to bleomycin challenge (7).

Proteins involved in cell cycle regulation and/or proliferation are septin-2, GTP-binding nuclear protein RAN, nucleoside diphosphate kinase A, and Rap1B (Figure 1A-D). Deviant regulation of these proteins has been described in association with several types of cancer (8-10). For example, septin-2 is a GTPase that plays a role in cytokinesis, its expression is cell-cycle dependent and is highest in G_2/M (11). The reduced levels found for the hypersensitive phenotype may relate to the damage resistant growth that has been described by Cloos *et al* (5).

Two of the significantly changing proteins are involved in DNA repair: Ku80 and high mobility group protein 2 (HMG2, Figure 1E-F). Ku80 forms a heterodimer with Ku70 that recognizes lesions in DNA and recruits the DNA-dependent protein kinase catalytic subunit (DNA-PK) to the sites of DNA damage, thereby starting DNA repair through non-homologous end joining (12). HMG2 binds to DNA, in particular to four-stranded structures that occur during DNA DSB repair (13) and it facilitates DNA binding of DNA-PK (14). In general, HMG proteins are involved in many DNA related processes, and besides the role in DNA repair, an important role has been found in protection of DNA against exogenous damaging agents (13,15). The patterns of HMG2 (Figure 1F) as well as of the heterogeneous nuclear ribonucleoprotein C1/C2 (hnRNP C1/C2) protein levels in response to DNA damage are remarkable, since they exhibit a significant interaction between phenotype and response in time. DNA damage-dependent chromatin binding was recently reported for hnRNP C1/C2 (16). Reduced Ku80, HMG2 and hnRNP C1/C2 levels in the hypersensitive phenotypes before and shortly after DNA damage induction may implicate a decrease in DNA protection and DNA damage repair. It is however not straightforward to relate the differential patterns of the proteins in Table 1 to the susceptibility to cancer development, since these proteins are each part of more than one complex with different functions in the DNA damage response. HMG2, for example, is part of the SET complex, the specific inhibitor of the tumor suppressor NM23-H1 (17). This complex comprises SET, acidic leucine-rich nuclear phosphoprotein 32A (pp32A), HMG2, Ape-1, and nucleoside diphosphate kinase A, of which some are differentially regulated in normal control versus hypersensitive cells (Table 1). The nuclear protein pp32A shows significantly higher levels for the hypersensitive phenotype (Figure 1G). Besides its role in the SET complex, this protein is also part of the tripartite INHAT complex that has been shown to play a role in transcription by regulating chromatin structure (17-19). Recently we suggested that this INHAT complex may have a similar function in DNA repair (7). The increased level detected for the hypersensitive phenotype might result in enhanced chromatin condensation, with the consequence that DNA repair enzymes have reduced access to the DNA lesions. This could also account for the enhanced mutagen susceptibility of this phenotype.

In conclusion, the comparative proteome analysis of lymphoblastoid nuclear extracts derived from cells exhibiting both mutagen hypersensitive and normal control phenotypes described here, has revealed a number of proteins with potential significance in understanding pathways involved in cancer susceptibility.

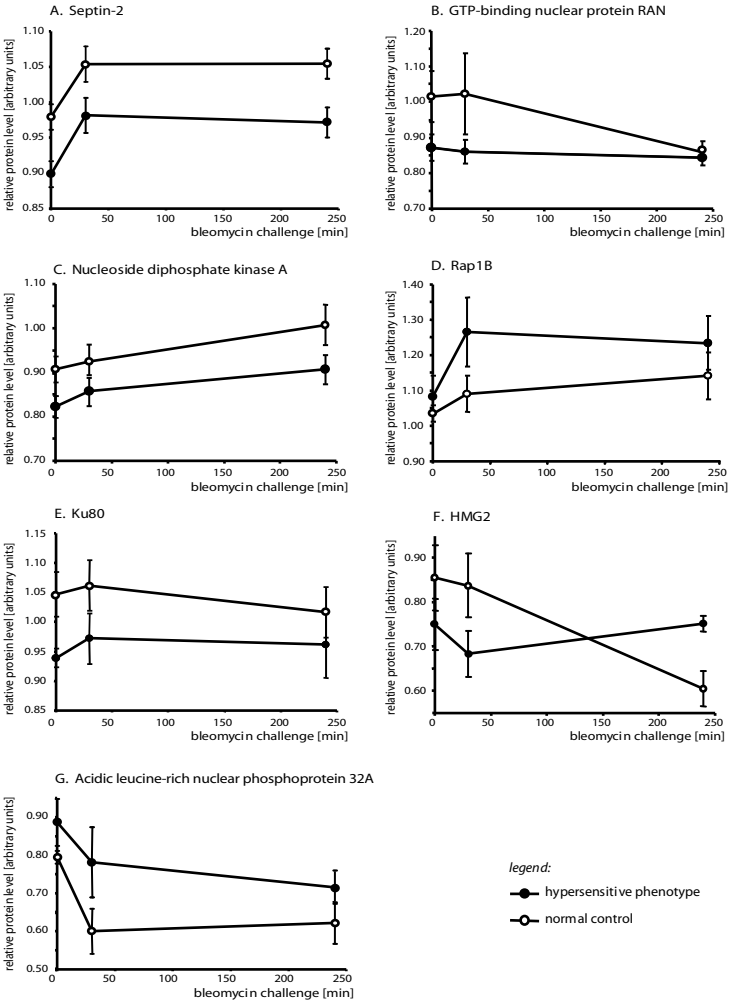


Figure 1. *Patterns of proteins that were differentially regulated between the hypersensitive phenotype and the normal control, this is the subset related to cell cycle regulation, cell progression and/or DNA repair. A) Septin-2; B) GTP-binding nuclear protein RAN; C) Nucleoside diphosphate kinase A; D) Rap1B; E) Ku80; F) HMG2; G) acidic leucine-rich nuclear phosphoprotein 32A. Each data point represents the average relative level of the protein in seven nuclear extracts, either for the normally sensitive (open dots) or for the hypersensitive phenotype (black dots); error bars correspond to the standard deviation of the mean.*

REFERENCES

1. Kastan, M. B., and Lim, D. S. (2000) The many substrates and functions of ATM. Nat Rev Mol Cell Biol **1**, 179-186
2. Cloos, J., Spitz, M. R., Schantz, S. P., Hsu, T. C., Zhang, Z. F., Tobi, H., Braakhuis, B. J., and Snow, G. B. (1996) Genetic susceptibility to head and neck squamous cell carcinoma. J Natl Cancer Inst **88**, 530-535
3. Cloos, J., Leemans, C. R., van der Sterre, M. L., Kuik, D. J., Snow, G. B., and Braakhuis, B. J. (2000) Mutagen sensitivity as a biomarker for second primary tumors after head and neck squamous cell carcinoma. Cancer Epidemiol Biomarkers Prev **9**, 713-717
4. Cloos, J., Nieuwenhuis, E. J., Boomsma, D. I., Kuik, D. J., van der Sterre, M. L., Arwert, F., Snow, G. B., and Braakhuis, B. J. (1999) Inherited susceptibility to bleomycin-induced chromatid breaks in cultured peripheral blood lymphocytes. J Natl Cancer Inst **91**, 1125-1130
5. Cloos, J., Temmink, O., Ceelen, M., Snel, M. H., Leemans, C. R., and Braakhuis, B. J. (2002) Involvement of cell cycle control in bleomycin-induced mutagen sensitivity. Environ Mol Mutagen **40**, 79-84
6. Cloos, J., de Boer, W. P., Snel, M. H., van den Ijssel, P., Ylstra, B., Leemans, C. R., Brakenhoff, R. H., and Braakhuis, B. J. (2006) Microarray analysis of bleomycin-exposed lymphoblastoid cells for identifying cancer susceptibility genes. Mol Cancer Res **4**, 71-77
7. Dirksen, E. H., Cloos, J., Braakhuis, B. J., Brakenhoff, R. H., Heck, A. J., and Slijper, M. (2006) Human lymphoblastoid proteome analysis reveals a role for the inhibitor of acetyltransferases complex in DNA double-strand break response. Cancer Res **66**, 1473-1480
8. Zhang, Z., Mitra, R. S., Henson, B. S., Datta, N. S., McCauley, L. K., Kumar, P., Lee, J. S., Carey, T. E., and D'Silva, N. J. (2006) Rap1GAP inhibits tumor growth in oropharyngeal squamous cell carcinoma. Am J Pathol **168**, 585-596
9. Mitra, R. S., Zhang, Z., Henson, B. S., Kurnit, D. M., Carey, T. E., and D'Silva, N. J. (2003) Rap1A and rap1B ras-family proteins are prominently expressed in the nucleus of squamous carcinomas: nuclear translocation of GTP-bound active form. Oncogene **22**, 6243-6256
10. Kim, D. S., Hubbard, S. L., Peraud, A., Salhia, B., Sakai, K., and Rutka, J. T. (2004) Analysis of mammalian septin expression in human malignant brain tumors. Neoplasia **6**, 168-178
11. Sakai, K., Kurimoto, M., Tsugu, A., Hubbard, S. L., Trimble, W. S., and Rutka, J. T. (2002) Expression of Nedd5, a mammalian septin, in human brain tumors. J Neurooncol **57**, 169-177
12. Jackson, S. P. (2002) Sensing and repairing DNA double-strand breaks. Carcinogenesis **23**, 687-696
13. Wunderlich, V., and Bottger, M. (1997) High-mobility-group proteins and cancer--an emerging link. J Cancer Res Clin Oncol **123**, 133-140
14. Watanabe, F., Shirakawa, H., Yoshida, M., Tsukada, K., and Teraoka, H. (1994) Stimulation of DNA-dependent protein kinase activity by high mobility group proteins 1 and 2. Biochem Biophys Res Commun **202**, 736-742
15. Pandita, T. K., and Hittelman, W. N. (1995) Evidence of a chromatin basis for increased mutagen sensitivity associated with multiple primary malignancies of the head and neck. Int J Cancer **61**, 738-743
16. Lee, S. Y., Park, J. H., Kim, S., Park, E. J., Yun, Y., and Kwon, J. (2005) A proteomics approach for the identification of nucleophosmin and heterogeneous nuclear ribonucleoprotein C1/C2 as chromatin-binding proteins in response to DNA double-strand breaks. Biochem J **388**, 7-15
17. Fan, Z., Beresford, P. J., Oh, D. Y., Zhang, D., and Lieberman, J. (2003) Tumor suppressor NM23-H1 is a granzyme A-activated DNase during CTL-mediated apoptosis, and the nucleosome assembly protein SET is its inhibitor. Cell **112**, 659-672
18. Seo, S. B., Macfarlan, T., McNamara, P., Hong, R., Mukai, Y., Heo, S., and Chakravarti, D. (2002) Regulation of histone acetylation and transcription by nuclear protein pp32, a subunit of the INHAT complex. J Biol Chem **277**, 14005-14010

19. Seo, S. B., McNamara, P., Heo, S., Turner, A., Lane, W. S., and Chakravarti, D. (2001) Regulation of histone acetylation and transcription by INHAT, a human cellular complex containing the set oncoprotein. Cell **104**, 119-130

Chapter 5

Development of a novel chemical probe for the selective enrichment of phosphorylated serine- and threonine-containing peptides

Eef H.C. Dirksen[*, 1], *Pieter van der Veken*[*, 1, 2], *Eelco Ruijter*[*, 1, 2], *Ronald C. Elgersma*[2], *Albert J.R. Heck*[1], *Dirk T.S. Rijkers*[2], *Monique Slijper*[1], *and Rob M.J. Liskamp*[2]

*These authors contributed equally to this work

[1] *Department of Biomolecular Mass Spectrometry, Utrecht Institute for Pharmaceutical Sciences and Bijvoet Centre for Molecular Research, Utrecht University, Utrecht, The Netherlands*

[2] *Department of Medicinal Chemistry, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands*

Based on: *ChemBioChem*, **2005**, 6, 2271-2280

ABSTRACT

Gaining insight into phosphoproteomes is of utmost importance for the understanding of regulation processes such as signal transduction and cellular differentiation. While the identification of phosphotyrosine-containing amino acid sequences in peptides and proteins is becoming possible, mainly because of the availability of specific, high-affinity antibodies, no general and robust methodology is presently available allowing the selective enrichment and analysis of serine- and threonine-phosphorylated proteins and peptides. The method presented here involves chemical modification of phosphorylated serine or threonine residues and subsequent derivatization using a new probe molecule. The designed probe consists of four parts: a reactive group that is used to specifically bind to the modified phosphopeptide, a part in which, optionally, stable isotopes can be incorporated, an acid-labile linker and an affinity tag for the selective enrichment of modified phosphopeptides from complex mixtures. The acid-cleavable linker allows full recovery from the affinity purified material and removal of the affinity tag prior to MS analysis. Next to the preparation of a representative probe molecule containing a biotin affinity tag, its applicability in phosphoproteome analysis is shown in a number of well-defined model systems with an increasing degree of complexity. During the development of the β-elimination/nucleophilic addition protocol special attention was paid to the different experimental parameters that may affect the chemical modification steps carried out on phosphorylated residues.

INTRODUCTION

The reversible phosphorylation of serine, threonine and tyrosine residues in peptides and proteins is generally recognized as a key event in the regulation of virtually all cellular functions. With an estimated one third of all eukaryotic proteins being amenable to this form of post-translational modification, phosphorylation has been well appreciated as a regulating mechanism touching almost every known signaling pathway (1). In spite of its importance, the global analysis of phosphoproteomes remains a challenge that is far from being realized. Problems related to heterogeneous phosphorylation patterns within a given protein and their relatively low abundance are often encountered, necessitating the development of selective enrichment techniques. In general, there is a clear need for robust analytical methodologies that allow the quantitative mapping of phosphoproteins, the sites of phosphorylation and the rapid screening for deviant phosphorylation events.

Several approaches towards quantitative phosphoproteomics have been described in literature (2-5), and as in many fields in proteomics, mass spectrometric analysis plays a crucial role in these studies (6-10). While analysis of phosphorylated tyrosine residues in peptides and proteins is facilitated by the commercial availability of specific antibodies (11-13), the use of immunochemical techniques for the enrichment and detection of serine- and threonine-phosphorylated proteins (accounting for >99% of all phosphorylated species) is rarely demonstrated (7,14). For tackling analytical problems associated with the latter two types of phosphorylated amino acid residues, different strategies based on affinity chromatographic enrichment of phosphopeptides from tryptic protein digests, followed by mass spectrometric analysis have been reported (15-22). Nevertheless, due to the poor ionization efficiency of phosphopeptides in the positive ionization mode, probably resulting from proton sequestration by the acidic phosphate group, methods in which the phosphate ester is chemically modified prior to MS can be expected to be more sensitive (9). To date, a number of methods for the analysis and/or enrichment of phosphorylated proteins and peptides based on chemical modification have been described (15,23-31). However, some of these approaches suffer from certain limitations, like relatively low efficiency of reactions. This results in low yields, hampering proteomics applications. We chose a stepwise development of a chemical modification/multifunctional probe addition strategy in attempts towards general methods for assessing phosphoserine/threonine containing (phospho)proteomes. During this process we focused on improvement of efficiency of the individual reactions and a stepwise extrapolation of the results obtained from the test conditions.

EXPERIMENTAL PROCEDURES

All solvents were distilled prior to use or were HPLC grade. Avidin immobilized on 6% agarose was purchased from Sigma. All other reagents were purchased from Fluka, Sigma-Aldrich, and Acros and used without further purification.

MALDI-TOF spectra were acquired in reflectron positive ion mode on an Applied Biosystems Voyager DE-STR or an Applied Biosystems 4700 Proteomics Analyzer using α-cyano-4-hydroxycinnamic acid as the matrix. MS/MS spectra were acquired on an Applied Biosystems 4700 Proteomics Analyzer.

Peptide synthesis

Tripeptides 5a and 5b were synthesized with standard *tert*-butyloxycarbonyl protection strategy using BOP as coupling agent and Et_3N as base in CH_2Cl_2 using N-Boc-*O*-benzylserine or –threonine.

Octapeptides 6a and 6b were synthesized on an Applied Biosystems 433A peptide synthesizer with (*p*-[(*R*,*S*)-α-[1-(9H-fluoren-9-yl)-methoxyformamido]-2,4-dimethoxy-benzyl]-phenoxyacetic acid)-crosslinked polystyrene as the starting resin. Phosphorylation was achieved by shaking the resin with reagent 11a (for 6a) or reagent 11b (for 6b). The crude peptide was precipitated from cold *tert*-butyl methyl ether and hexane (50/50), redissolved in H_2O/*tert*-butanol (50/50) and lyophilized.

Synthesis of probe molecules

The synthesis of probe molecules used in this study was as described elsewhere (32).

Modification of phosphopeptides

A mixture of peptides (10 μl, containing 0.1-1 nmol of phosphopeptide in 10-50% aq. MeOH), MeOH (80 μl), and saturated aqueous $Ba(OH)_2$ (3 μl, 0.15 N, 450 nmol) in a was incubated in an Eppendorf Thermomixer apparatus (800 rpm) for 3 h at 45°C. Then, 0.30 N aq. H_2SO_4 was added (1.5 μl, 450 nmol), and the mixture was vortexed and centrifugated (4 min. at 12000 rpm) to precipitate $BaSO_4$ and $Ba_3(PO_4)_2$. The supernatant was transferred to a new reaction vial and EDT (15 μl) and Et_3N (15 μl) were added. The resulting mixture was vortexed and shaken (800 rpm) for 48 h at 45°C. An aqueous solution of TCEP·HCl (Tris(carboxyethyl)phosphine HCl salt, 10 μl, 1 mM, 10 nmol) was added and the vial was vortexed and briefly centrifugated again. All volatile components were evaporated at 45°C with gentle shaking (300 rpm) under a continuous argon stream. MeOH (100 μl) was added to the residue, and the vial was vortexed and centrifugated (3 min. at 12000 rpm). After evaporation of volatile components, an additional 100 μl of MeOH was added and evaporated again; this was repeated once. The residue was dissolved in MeOH (90 μl) and a 5 mM solution of 9 in MeOH (10 μl, 50 nmol) and Et_3N (100 mM in MeOH, 10 μl) was added. The mixture was vortexed and shaken (800 rpm) for 3 h at 45°C. Unreacted maleimide groups were capped by addition of β-mercaptoethanol (1 μl). Again, volatile components were evaporated and the residue was redissolved in 100 mM phosphate buffer (pH 7.4) containing 200 mM NaCl (30 μl). Subsequently, 10 μl of a slurry of avidin immobilized on 6% agarose was added. The mixture was gently shaken for 2 h (1000 rpm) at 25°C. The supernatant was removed and the beads were washed

with 5 x 30 μl phosphate buffer. Subsequently the beads were incubated (1000 rpm, 25°C) for 1 h with 95% TFA (30 μl) and the supernatant containing the modified peptide(s) was transferred to a reaction vial. After final evaporation of volatile components the residue was then analyzed by MALDI-TOF mass spectrometry.

RESULTS AND DISCUSSION

Our approach for studying phosphorylated serine and threonine residues in peptides and proteins exploits the well-described base-promoted β-elimination reaction of aliphatic phosphate esters (1) in basic media (24,25,29,33), as illustrated in Figure 1A. In this step, a Michael acceptor moiety 2 is created which is susceptible to attack by nucleophiles, such as thiols. We therefore developed probe molecule 4, of which the structure is given in Figure 1B. The probe consists of four functional modules: a nucleophilic reactive site (thiol), a part in which heavy isotopes can be built in (ethylene chain), an acid-labile "Wang"-type linker, and a biotin moiety that allows affinity-based purification.

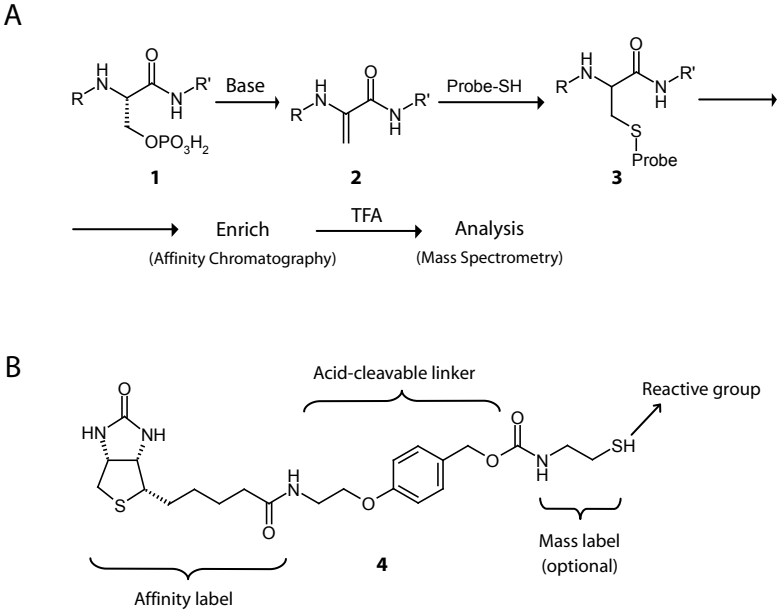
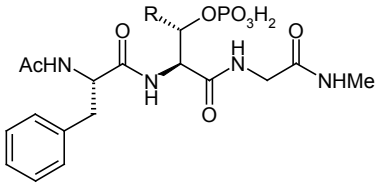
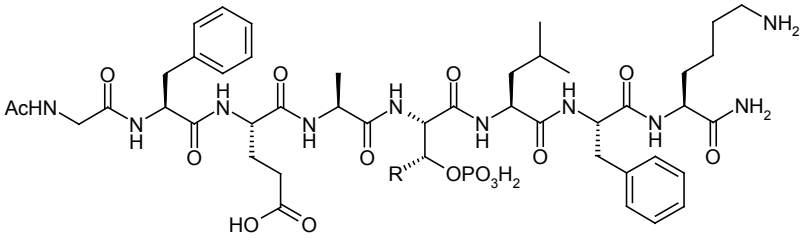


Figure 1. *(A) Base induced β-elimination followed by Michael addition of a SH-containing probe. (B) The thiol-containing probe.*

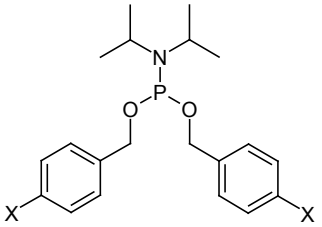
For evaluation of our labeling protocol and analysis of phosphopeptides, the following test systems -with an increasing degree of complexity- were developed (see Figure 2): phosphorylated tripeptides 5a and 5b, phosphorylated octapeptides 6a and 6b, and a mixture of peptide 6a and a number of other, nonphosphorylated, peptides. This mixture was considered a model for protein tryptic digests, and served to assess affinity enrichment after probe attachment.



5a: R = H, Ac-F**pS**G-NH_2
5b: R = Me, Ac-F**pT**G-NH_2



6a: R = H, Ac-GFEA**pS**LFK-NH_2
6b: R = Me, Ac-GFEA**pT**LFK-NH_2



11a = Cl
11b = OMe

Figure 2. *Model tri- and octapeptides*

Incubation of 4 with β-eliminated peptides derived from *e.g.* 6a or 6b was expected to lead to the formation of a covalent adduct by Michael addition of the sulfhydryl moiety of 4 to the α, β-unsaturated amino acid residue in the peptide (as shown in Figure 1A). However, no covalent linkage of the probe to the peptide was observed, even upon addition of a tenfold excess of 4 to a solution of a β-eliminated peptide derived from 6a. The size of the nucleophile is probably detrimental for its reactivity and a smaller thiol nucleophile -to be used in high excess- might be more advantageous. This enticed us to consider alternative strategies.

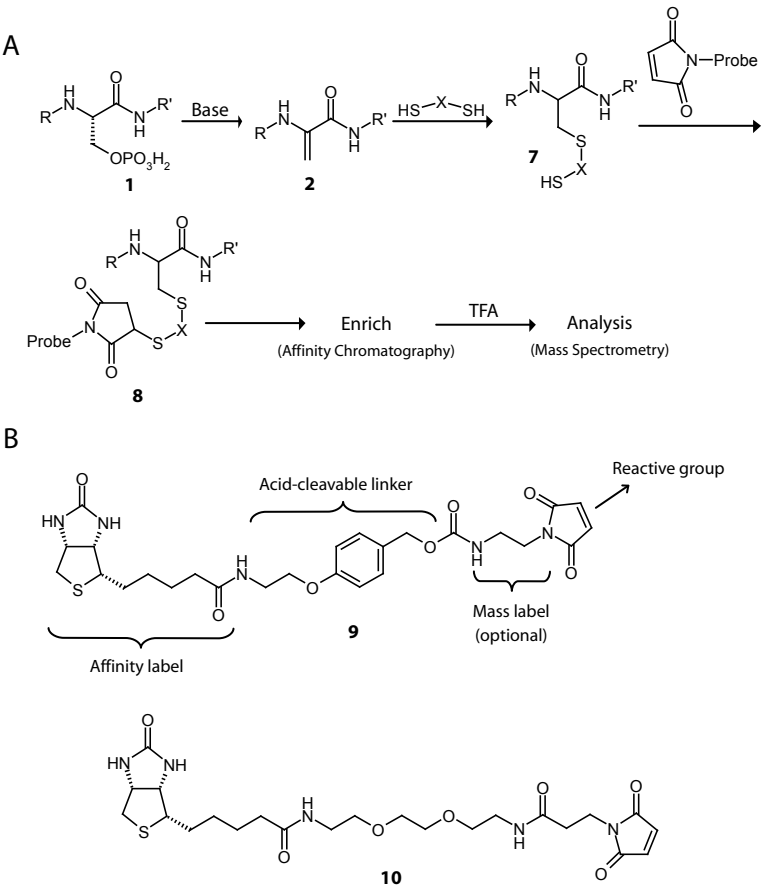


Figure 3. *Dithiol modification of phosphopeptides and electrophilic addition of the probe.*

A possible strategy, exemplified in Figure 3A, involves the addition of a dithiol to β-eliminated phosphopeptides 2. The formerly phosphorylated serine or threonine residue is thereby functionalized with a thiol group (7), which serves as a handle for linking the peptide to a probe molecule. A similar successful approach has been reported (25,27,29,34). However, the conditions described in the literature were not successful in our hands for modification of phosphopeptides 6a and 6b. Therefore, an optimized procedure for β-elimination and subsequent ethanedithiol (EDT) addition was developed. In addition to supplying a handle for probe attachment, the generation of a thiol-functionalized residue allows the application of available methods for cysteine labeling. So in comparison with 4, the new probe carried a maleimide functionality, which is an electrophilic reactive group that can be coupled to the sulfhydryl group on the formerly phosphorylated peptide 7. Compound 9 of which the structure is given in Figure 3B, again included a biotin affinity label, Wang-type linker, and a maleimide functionality.

A similar, commercially available, probe that lacks the acid-sensitive part (10, shown in Figure 3B) has been used earlier by Oda *et al* for isolating a phosphorylated protein that was spiked into a complex mixture (27). Other methods derived from this strategy have also been described (34,35). However, the presence of the cleavable linker in our molecule offers means to circumvent several of the shortcomings of the technique, such as incomplete recovery from the (strept)avidin material used during the affinity-based isolation/enrichment. Next to that, incorporation of the acid-labile linker in 9 affords small, stable adducts - that no longer include the biotin moiety - on formerly phosphorylated serine and threonine residues, leading to more easy interpretable mass spectra.

The validity of these assumptions was tested using well-defined model systems. For this, two model octapeptides, 6a and 6b, containing either a phosphorylated serine or a phosphorylated threonine residue, were used in experiments in which the protocol of β-elimination/Michael addition/probe addition and avidin affinity chromatography was applied to isolate and enrich both peptides from peptide mixtures with known compositions. In addition, since no systematic optimization studies of the β-elimination and Michael addition reactions could be found in literature and existing reports in general exclusively deal with the reactivity of serine-phosphorylated peptides, two phosphorylated model tripeptides, 5a and 5b, were prepared with the aim of studying experimental parameters affecting these transformations for both serine- and threonine-phosphorylated peptides. This was done on a preparative scale (10-100 mg), allowing a thorough characterization of all products formed.

For the optimization of the β-elimination reaction, peptides were incubated at 45°C with different inorganic (NaOH, CsOH, $Ba(OH)_2$) and organic bases (DBU, $NaOCH_3$, Et_3N) in $MeCN/H_2O$ mixtures (60:40 to 90:10 v/v). For every solvent mixture tested, experiments were run with two, five seven and ten molar equivalents of base. In all cases, nearly quantitative elimination of phosphate from the serine-phosphorylated peptide could be achieved within 3 hours. Formation of the dehydrobutyrine residue in the phosphothreonine containing peptide however turned out to be significantly slower, requiring up to 6 hours of incubation: hydrolysis of the phosphate ester could be detected for the alkali bases and was most pronounced in experiments with NaOH. The product of this competing side reaction, Ac-FTG-NHMe, could be isolated in yields up to 10%. Although none of the used bases was found to be a significantly superior promoter of the β-elimination reaction, we decided to use $Ba(OH)_2$ in all further experiments. Due to the production of highly insoluble barium phosphate, the latter has been suggested to promote the β-elimination of phosphate esters with higher efficiency compared to the analogous elimination of *O*-linked sugar residues, thereby minimizing the risk of forming dehydro-amino acid residues in glycopeptides present in complex biological samples (36).

Although several authors claim that the use of elimination mixtures containing the highly polar aprotic DMSO gave favorable results in terms of the kinetics and yields of the reaction (29,34), this observation could not be reproduced with our model tripeptides. Results obtained when the elimination reaction was run in $MeCN/H_2O$ or in an elimination mixture containing DMSO, EtOH and H_2O (3:1:1) were comparable. Furthermore, due to the poor compatibility of DMSO with the process of matrix crystal formation (MALDI) and electrospray ionization, solvent systems based on $MeCN/H_2O$ mixtures were considered to be more suitable for this application.

After isolation of the dephosphorylated peptides, optimization of the dithiol addition reaction was performed by incubating with two, four or six equivalents of 1,4-dithiothreitol (DTT) and a catalytic amount of base (Et_3N, 0.2 eq.). In all cases, DTT addition proceeded in a quantitative fashion within 3 hours. Again, no additional advantages in terms of yields and kinetics were observed with the $DMSO/EtOH/H_2O$ mixture as compared to $MeCN/H_2O$ systems.

Subsequently, experiments were carried out in order to verify if the phosphopeptide in amounts of 10 pmol to 10 nmol could be isolated from solution using the series of transformations described earlier. Conditions optimized for the β-elimination and dithiol addition of model tripeptides were applied: $MeCN/H_2O$ (90:10, v/v) was chosen as the solvent and $Ba(OH)_2$ as the base. The

process of β-elimination was found to be similarly straightforward but invariantly accompanied by some hydrolysis of the phosphate ester, both for the serine- and threonine-phosphorylated peptide. Addition of DTT however, proved to become substantially more difficult on downscaling: even with 50 equivalents, only a minor fraction of the dehydropeptides could be transformed into the corresponding dithiol adducts. Since the introduction of even larger excess of DTT was expected to compromise the subsequent process of probe binding due to scavenging of maleimide by unreacted dithiol, we decided to use volatile ethanedithiol (EDT) in all further experiments. Prior to probe addition, unreacted portions of the latter could then be removed from the system by evaporation. However, a large excess (~10^4 eq.) of EDT (as well as of Et_3N) was required to achieve near-quantitative dithiol addition. This required excess led to the formation of unexpected side products when MeCN was used as solvent: next to the expected peaks in the MS spectrum of the dehydropeptide and the EDT adduct, several other peaks were observed, which all displayed a mass increase of 41 Dalton with respect to the expected peaks. MS/MS analysis revealed that these peaks arose from addition of MeCN to the lysine residue. We suspect that high concentrations of EDT and base in MeCN lead to the formation of a thioacetimidate, which are susceptible to attack by nucleophiles such as the lysine residue epsilon amino group.

This observation led us to change the solvent of EDT addition (and β-elimination) to MeOH. Although this requires somewhat longer reaction times, β-elimination and EDT addition proceeded smoothly, and no unexpected side products were observed.

The phosphorylated peptides 6a and 6b were subjected to β-elimination/EDT addition protocol and subsequently affinity purified using immobilized avidin. After cleavage of the Wang-type linker, the supernatant was analyzed using MALDI-TOF mass spectrometry. This showed that the former phosphorylated serine and threonine-containing peptides had been modified. The subsequent mass spectrometric fragmentation of the modified peptides is shown in Figure 4. The resulting MS/MS spectrum not only contains nearly complete series of *b* and *y* ions, that can be used for sequencing of the peptide and determination of the site of phosphorylation, but also shows the immonium ion of the modified residues at *m/z* 276.1 (in case of the serine-modified peptide) and *m/z* 290.1 (in case of the threonine-modified peptide) that show that indeed, the phosphorylated amino acid has been affected in the desired fashion, by the protocol used. These specific immonium ions may potentially be used as marker ions for the phosphorylated serine and threonine peptides in mass spectrometric analyses.

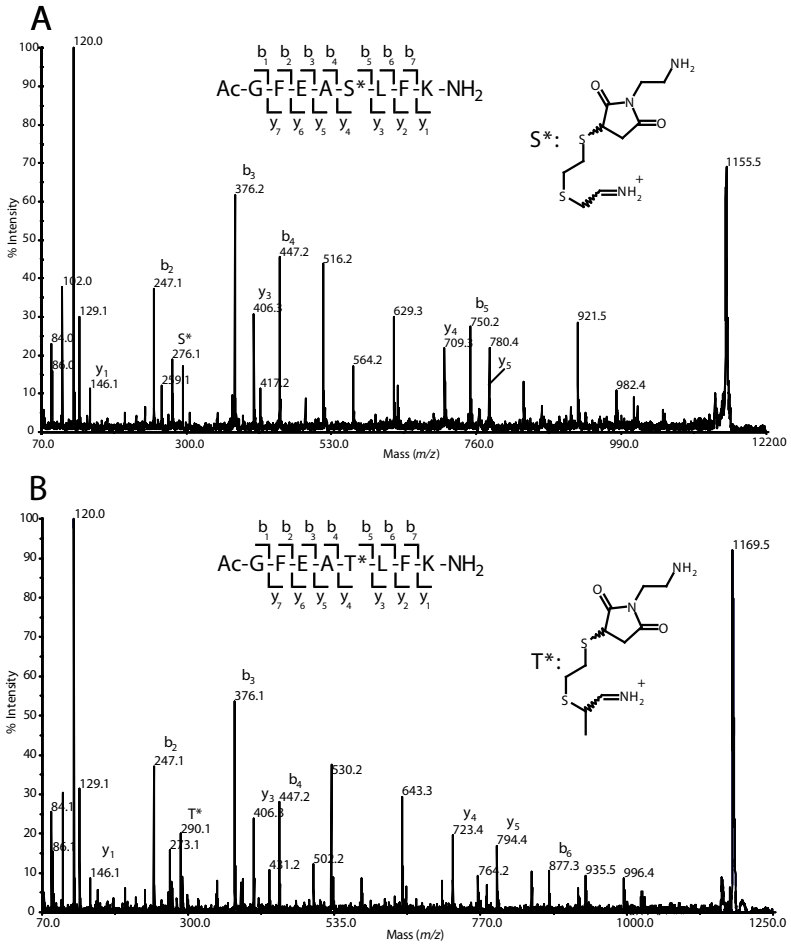


Figure 4. *Mass spectrometric fragmentation (MS/MS) spectra of the modified phosphopeptides 6a and 6b. Characteristic b and y type ions that were used for sequencing of the peptide are indicated. A) MS/MS spectrum of the modified serine-phosphorylated peptide (m/z 1155.5). The immonium ion of the modified serine residue at m/z 276.1, of which the structure is given in the inset, is marked S*. The ion at m/z 921.5 corresponds to the peptide that has lost the complete modification as a result of β-elimination during MS/MS. The ions at m/z 629.3 and 516.2 correspond to the b_6 and b_5 ions of the peptide that underwent β-elimination. B) MS/MS spectrum of the modified threonine-phosphorylated peptide (m/z 1169.5). The immonium ion of the modified serine residue at m/z 290.1, of which the structure is given in the inset, is marked T*. The ion at m/z 935.5 corresponds to the peptide that has lost the complete modification as a result of β-elimination during MS/MS. The ions at m/z 643.3 and 530.2 correspond to the b_6 and b_5 ions of the peptide that underwent β-elimination.*

Compared to fragmentation of a phosphorylated peptide, that normally mainly results in loss of the labile phosphate group and requires further (MS/MS/MS) fragmentation to gain sequence information, the approach shown here has the advantage that the majority of the label stays on the amino acid residue, thereby allowing sequencing and phosphorylation site determination. In addition, the modification results in the introduction of an extra amine group, which increases the ionization efficiency of the peptide, especially when compared to the phosphorylated peptide.

Finally, to confirm the applicability of our enrichment strategy, we composed a mixture of unpurified, synthetic non-phosphorylated peptides varying in amino acid composition and ranging in molecular weight from 900 to 3300 Da, to which the serine phosphorylated octapeptide 6a was added (~1 nmol each, see Figure 5A).

The peptide mixture was subjected to the optimized β-elimination/EDT addition protocol. After evaporation of the volatiles, the residue was dissolved in MeOH (90 μL) and incubated at 45°C for 2 hours with probe molecule 9 (5-50 eq). Subsequently, the mixture was incubated with immobilized avidin for 1 hour. Then the beads were washed with buffer and treated with TFA to cleave the Wang-type linker. After removal of TFA, the mixture was analyzed using MALDI-TOF mass spectrometry. The mass spectrum of the enriched fraction is shown in Figure 5B. The most abundant peak at *m/z* 1155.5 corresponds to the modified phosphopeptide. The presence of other compounds (peptides) in the enriched fraction might be caused by aspecific binding of particular peptides to either avidin or agarose on which the avidin was immobilized. These peaks however, can be used to calculate a semi-quantitative amount of enrichment: the intensity ratio between the synthetic peptide with an *m/z* value of 1127.6 and the phosphopeptide (*m/z* 1019.4, marked with P in Figure 5A) before enrichment is 10, while it is 0.11 after enrichment, yielding a semi-quantitative enrichment of approximately 100-fold.

Together with the sequencing and identification of the former site of phosphorylation, as shown in Figure 4, these results confirm the validity of our methodology for modification, enrichment and analysis of serine- and threonine-phosphorylated peptides.

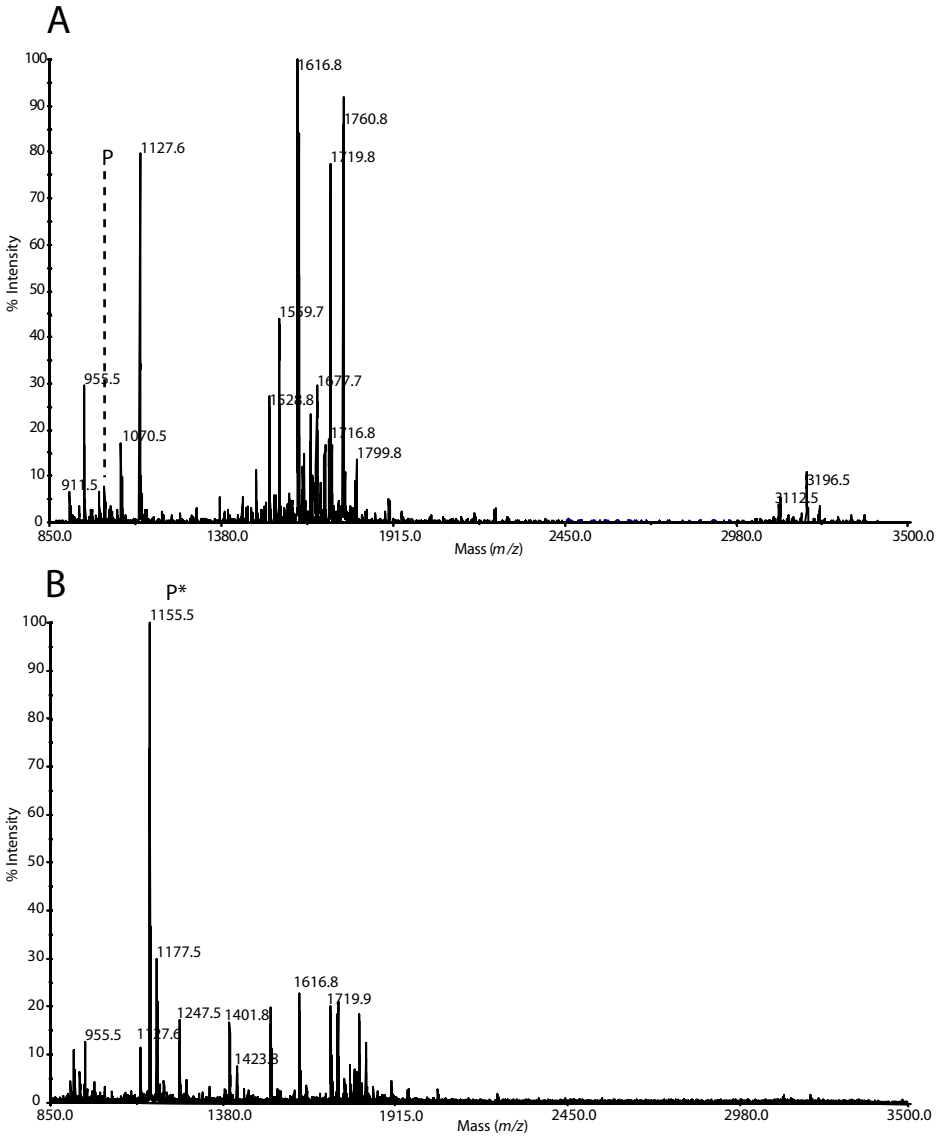


Figure 5. *A) MALDI-TOF mass spectrum of the starting mixture of peptides that was used to mimic a tryptic digest to test the chemical modification procedure. The phosphopeptide (6a) at m/z 1019.4 is indicated by a dashed line and the letter P. B) MALDI-TOF mass spectrum of the supernatant obtained after affinity enrichment and cleavage of the Wang-type linker. The modified phosphopeptide is indicated by P* ($[M+H]^+$, m/z 1155.5, $[M+Na]^+$, m/z 1177.5).*

CONCLUSIONS

The analysis of post-translational modifications on proteins in complex mixtures, especially protein phosphorylation, requires methods that allow specific enrichment of peptides that carry this type of modification. This is on the one hand due to relatively low abundance and phosphorylation stochiometry and on the other due to the fact that mass spectrometry, routinely used for these purposes, has some disadvantages that hamper the investigation of this modification. The extra negative charge for example, lowers the ionization efficiency of phosphorylated peptides in the positive ion mode. In addition, MS/MS collision induced fragmentation of these ions mainly results in loss of the labile phosphate group, obstructing further sequence analysis.

In order to improve the analysis of phosphorylated peptides we have developed a protocol for the selective β-elimination and subsequent EDT modification of serine- and threonine-phosphorylated peptides using three different test systems with different degrees of complexity. In our test systems, this newly developed protocol proved more effective than previously described methods. In addition, we have developed a multifunctional molecular probe featuring a biotin affinity label, an acid-labile linker, a site for the optional incorporation of stable isotopes, and an electrophilic reactive group. This probe molecule selectively allows further modification of previously phosphorylated serine and threonine residues, as well as affinity-based purification of modified phosphopeptides from a mixture of peptides. Modified peptides can be released from the affinity purification material by cleavage of the acid-labile linker, and the remaining, compact adduct allows identification of the peptide sequence, as well as of the phosphorylation site. Incorporation of an acid-labile linker ensures quantitative release of bound peptides from the affinity purification material in comparison with previously reported biotin-based molecular probes. In addition, MS/MS spectra of the modified peptides are more straightforward when the adducts do not contain biotin, and sequencing of the peptide and assignment of the phosphorylation site is simpler.

The work described here offers an approach towards better analysis of phosphorylated peptides. It not only enables affinity enrichment of former phosphorylated peptides, but it also improves further analysis of these peptides by introduction of a stable group, that on top of that shows improved ionization due to an additional amino group.

It is clear, however, that the application of this method in biological (proteomic) samples, such as complex cellular lysates containing low amounts of phosphorylated proteins within the total protein pool, is not yet possible. The limit of detection of the method presented here is in the order of 1 nanomole.

This is mainly due to the fact that downscaling of chemical modification methods comprising multiple reaction steps is not straightforward. One aspect at which improvement of the detection limits can be realized, apart from optimizing the down-scaling of reactions is, for example, the use of more efficient (bio)affinity couples.

The introduction of stable isotopes, already briefly discussed here, can open up the path towards quantitative phosphoproteomic studies that not only give insight into the pool of phosphorylated proteins, but also reveal some of the dynamics of this modification in time or between different cellular states.

Taken together, gaining knowledge on down-scaling of chemical reactions and incorporation of stable isotopes for relative quantitation of phosphorylation levels provide sufficient challenges to further improve the field of chemistry-based phosphoproteomics and thereby allow the analysis of phosphorylated peptides in complex biological samples.

REFERENCES

1. Hunter, T. (2000) Signaling--2000 and beyond. *Cell* **100**, 113-127
2. Adam, G. C., Sorensen, E. J., and Cravatt, B. F. (2002) Chemical strategies for functional proteomics. *Mol Cell Proteomics* **1**, 781-790
3. Ahn, N. G., and Resing, K. A. (2001) Toward the phosphoproteome. *Nat Biotechnol* **19**, 317-318
4. Kalume, D. E., Molina, H., and Pandey, A. (2003) Tackling the phosphoproteome: tools and strategies. *Curr Opin Chem Biol* **7**, 64-69
5. Sechi, S., and Oda, Y. (2003) Quantitative proteomics using mass spectrometry. *Curr Opin Chem Biol* **7**, 70-77
6. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422**, 198-207
7. Gronborg, M., Kristiansen, T. Z., Stensballe, A., Andersen, J. S., Ohara, O., Mann, M., Jensen, O. N., and Pandey, A. (2002) A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate. *Mol Cell Proteomics* **1**, 517-527
8. Bennett, K. L., Stensballe, A., Podtelejnikov, A. V., Moniatte, M., and Jensen, O. N. (2002) Phosphopeptide detection and sequencing by matrix-assisted laser desorption/ionization quadrupole time-of-flight tandem mass spectrometry. *J Mass Spectrom* **37**, 179-190
9. Schweppe, R. E., Haydon, C. E., Lewis, T. S., Resing, K. A., and Ahn, N. G. (2003) The characterization of protein post-translational modifications by mass spectrometry. *Acc Chem Res* **36**, 453-461
10. Tao, W. A., and Aebersold, R. (2003) Advances in quantitative proteomics via stable isotope tagging and mass spectrometry. *Curr Opin Biotechnol* **14**, 110-118
11. Conrads, T. P., and Veenstra, T. D. (2005) An enriched look at tyrosine phosphorylation. *Nat Biotechnol* **23**, 36-37
12. Machida, K., Mayer, B. J., and Nollau, P. (2003) Profiling the global tyrosine phosphorylation state. *Mol Cell Proteomics* **2**, 215-233
13. Rush, J., Moritz, A., Lee, K. A., Guo, A., Goss, V. L., Spek, E. J., Zhang, H., Zha, X. M., Polakiewicz, R. D., and Comb, M. J. (2005) Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat Biotechnol* **23**, 94-101
14. Kane, S., Sano, H., Liu, S. C., Asara, J. M., Lane, W. S., Garner, C. C., and Lienhard, G. E. (2002) A Method to Identify Serine Kinase Substrates. Akt PHOSPHORYLATES A NOVEL ADIPOCYTE PROTEIN WITH A Rab GTPASE-ACTIVATING PROTEIN (GAP) DOMAIN. *J Biol Chem* **277**, 22115-22118
15. Ahn, Y. H., Park, E. J., Cho, K., Kim, J. Y., Ha, S. H., Ryu, S. H., and Yoo, J. S. (2004) Dynamic identification of phosphopeptides using immobilized metal ion affinity chromatography enrichment, subsequent partial beta-elimination/chemical tagging and matrix-assisted laser desorption/ionization mass spectrometric analysis. *Rapid Commun Mass Spectrom* **18**, 2495-2501
16. Cao, P., and Stults, J. T. (2000) Mapping the phosphorylation sites of proteins using on-line immobilized metal affinity chromatography/capillary electrophoresis/electrospray ionization multiple stage tandem mass spectrometry. *Rapid Commun Mass Spectrom* **14**, 1600-1606
17. Ficarro, S. B., McCleland, M. L., Stukenberg, P. T., Burke, D. J., Ross, M. M., Shabanowitz, J., Hunt, D. F., and White, F. M. (2002) Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat Biotechnol* **20**, 301-305
18. Metodiev, M. V., Timanova, A., and Stone, D. E. (2004) Differential phosphoproteome profiling by affinity capture and tandem matrix-assisted laser desorption/ionization mass spectrometry. *Proteomics* **4**, 1433-1438
19. Raska, C. S., Parker, C. E., Dominski, Z., Marzluff, W. F., Glish, G. L., Pope, R. M., and Borchers, C. H. (2002) Direct MALDI-MS/MS of phosphopeptides affinity-bound to immobilized metal ion affinity chromatography beads. *Anal Chem* **74**, 3429-3433

20. Pinkse, M. W., Uitto, P. M., Hilhorst, M. J., Ooms, B., and Heck, A. J. (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem* **76**, 3935-3943
21. Brill, L. M., Salomon, A. R., Ficarro, S. B., Mukherji, M., Stettler-Gill, M., and Peters, E. C. (2004) Robust phosphoproteomic profiling of tyrosine phosphorylation sites from human T cells using immobilized metal affinity chromatography and tandem mass spectrometry. *Anal Chem* **76**, 2763-2772
22. Nuhse, T. S., Stensballe, A., Jensen, O. N., and Peck, S. C. (2003) Large-scale Analysis of in Vivo Phosphorylated Membrane Proteins by Immobilized Metal Ion Affinity Chromatography and Mass Spectrometry. *Mol Cell Proteomics* **2**, 1234-1243
23. Adamczyk, M., Gebler, J. C., and Wu, J. (2001) Selective analysis of phosphopeptides within a protein mixture by chemical modification, reversible biotinylation and mass spectrometry. *Rapid Commun Mass Spectrom* **15**, 1481-1488
24. Jaffe, H., Veeranna, and Pant, H. C. (1998) Characterization of serine and threonine phosphorylation sites in beta-elimination/ethanethiol addition-modified proteins by electrospray tandem mass spectrometry and database searching. *Biochemistry* **37**, 16211-16224
25. McLachlin, D. T., and Chait, B. T. (2003) Improved beta-elimination-based affinity purification strategy for enrichment of phosphopeptides. *Anal Chem* **75**, 6826-6836
26. Molloy, M. P., and Andrews, P. C. (2001) Phosphopeptide derivatization signatures to identify serine and threonine phosphorylated peptides by mass spectrometry. *Anal Chem* **73**, 5387-5394
27. Oda, Y., Nagasu, T., and Chait, B. T. (2001) Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat Biotechnol* **19**, 379-382
28. Knight, Z. A., Schilling, B., Row, R. H., Kenski, D. M., Gibson, B. W., and Shokat, K. M. (2003) Phosphospecific proteolysis for mapping sites of protein phosphorylation. *Nat Biotechnol* **21**, 1047-1054
29. Thaler, F., Valsasina, B., Baldi, R., Xie, J., Stewart, A., Isacchi, A., Kalisz, H. M., and Rusconi, L. (2003) A new approach to phosphoserine and phosphothreonine analysis in peptides and proteins: chemical modification, enrichment via solid-phase reversible binding, and analysis by mass spectrometry. *Anal Bioanal Chem* **376**, 366-373
30. Vosseller, K., Hansen, K. C., Chalkley, R. J., Trinidad, J. C., Wells, L., Hart, G. W., and Burlingame, A. L. (2005) Quantitative analysis of both protein expression and serine / threonine post-translational modifications through stable isotope labeling with dithiothreitol. *Proteomics* **5**, 388-398
31. Zhou, H., Watts, J. D., and Aebersold, R. (2001) A systematic approach to the analysis of protein phosphorylation. *Nat Biotechnol* **19**, 375-378
32. van der Veken, P., Dirksen, E. H., Ruijter, E., Elgersma, R. C., Heck, A. J., Rijkers, D. T., Slijper, M., and Liskamp, R. M. (2005) Development of a Novel Chemical Probe for the Selective Enrichment of Phosphorylated Serine- and Threonine-Containing Peptides. *Chembiochem* **6** 2271-2280
33. Li, W., Boykins, R. A., Backlund, P. S., Wang, G., and Chen, H. C. (2002) Identification of phosphoserine and phosphothreonine as cysteic acid and beta-methylcysteic acid residues in peptides by tandem mass spectrometric sequencing. *Anal Chem* **74**, 5701-5710
34. Goshe, M. B., Conrads, T. P., Panisko, E. A., Angell, N. H., Veenstra, T. D., and Smith, R. D. (2001) Phosphoprotein isotope-coded affinity tag approach for isolating and quantitating phosphopeptides in proteome-wide analyses. *Anal Chem* **73**, 2578-2586
35. Weckwerth, W., Willmitzer, L., and Fiehn, O. (2000) Comparative quantification and identification of phosphoproteins using stable isotope labeling and liquid chromatography/mass spectrometry. *Rapid Commun Mass Spectrom* **14**, 1677-1681
36. Byford, M. F. (1991) Rapid and selective modification of phosphoserine residues catalysed by Ba2+ ions for their detection during peptide microsequencing. *Biochem J* **280** (Pt 1), 261-265

Chapter 6

Post-translational modification of histone chaperones in the early response to DNA damage

Eef H.C. Dirksen[1], Martijn W.H. Pinkse[1], Dirk T.S. Rijkers[2], Jacqueline Cloos[3], Rob M.J. Liskamp[2], Monique Slijper[1], Albert J.R. Heck[1]

[1] Department of Biomolecular Mass Spectrometry, Utrecht Institute for Pharmaceutical Sciences, Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands

[2] Department of Medicinal Chemistry, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands

[3] Section Tumor Biology, Department of Otolaryngology/Head-Neck Surgery, VU University Medical Center, Amsterdam, The Netherlands.

Based on: *Journal of Proteome Research*, ***2006***, in press

ABSTRACT

The cellular response to DNA double strand breaks is a multifaceted mechanism that involves detection of lesions in DNA and the subsequent repair to prevent genomic aberrations. Chromatin relaxation is required to provide access for DNA repair factors and to aid the reformation of chromatin structure following repair. For this, several proteins and protein complexes involved in chromatin remodelling events are recruited to core histone N-terminal tails. In an affinity-based proteomic approach, using a core histone consensus N-terminal tail (NTT) peptide as bait, we investigated the changes in interactions of nuclear proteins before and shortly after DNA double strand break induction. Over 700 identified proteins were found to bind specifically to the NTT, which makes it the most comprehensive proteomic survey of the broad spectrum of nuclear proteins interacting with the NTT of core histones in nucleosomes. The abundance of the majority of these proteins was found to be unchanged following DNA damage induction. Therefore, we focused our analysis on potential post-translational activation by (de)phosphorylation. An in-depth analysis of protein phosphorylation (more than 90 unique sites in over 60 proteins) revealed that the phosphorylation status of several proteins involved in chromatin remodelling, such as chromodomain-helicase-DNA-binding protein 1 (CHD1), structure-specific recognition protein 1 (SSRP1) and nucleosome assembly protein 1-like 4 (NAP-2), changes upon DNA damage induction. The chaperones that were found to be differentially phosphorylated upon DNA damage have in common that they are part of closely interacting protein complexes involved in regulatory mechanisms at the crossroads of nucleosome assembly, DNA replication, transcription and DNA repair.

INTRODUCTION

In eukaryotic cells, DNA is organized in chromatin, which is a dynamic higher-order structure that can adopt markedly different conformations (1). Most of the time chromatin, consisting of nucleosomes around which DNA is wrapped, is densely packed. This dense macro structure, also referred to as 'beads on a string', is stabilized by linker histones and a variety of other factors. The nucleosomes, which can be considered as the basic unit of chromatin, are built up of four pairs of core histones (two histone H2A-H2B and two histone H3-H4 pairs (2)), schematically shown in Figure 1C. The histone N-terminal tail (NTT) contains several basic (lysine and arginine) amino acids and protrudes from the nucleosome where it can interact both with negatively charged DNA and with proteins involved in DNA-related processes. A well-known feature of the amino

acid residues in the histone NTT is that they can undergo extensive post-translational modifications (3,4). These modifications on the one hand diminish interactions between the tails and DNA, thereby loosening chromatin structure (5). On the other hand they regulate interactions with histone chaperones, which are proteins characterized by negatively charged domains that can disturb DNA-histone contacts and thereby function in processes involving DNA. A 'compact' chromatin structure, i.e. heterochromatin, generally inhibits the binding of transcriptional regulators.

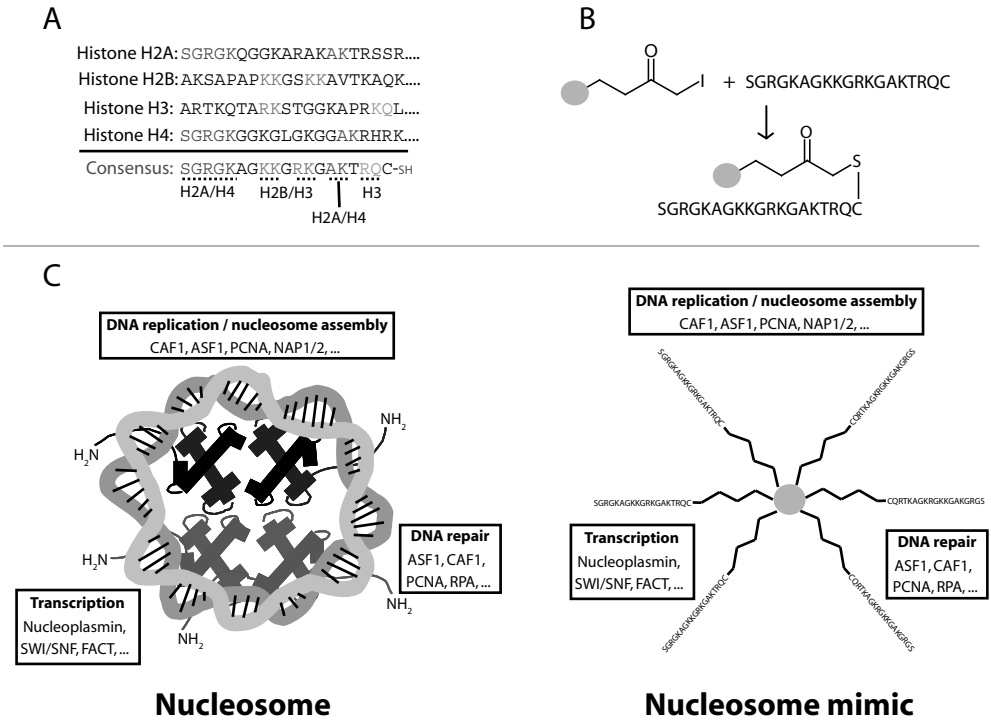


Figure 1. *(A) Design of the core histone N-terminal consensus peptide used for the affinity pull-downs. It contains various conserved sequence elements from the individual core histone N-terminal tails of H2A, H2B, H3 and H4, as indicated by the coloured residues. (B) The peptide was immobilized on iodoacetyl-functionalized agarose beads using the C-terminal cysteine. (C) Schematic overview of core histones in a nucleosome and examples of histone-interacting proteins involved in nucleosome assembly, transcription and DNA repair (left panel). The experimental set up (nucleosome mimic) used to enrich for a broad range of histone chaperones using the immobilized core histone consensus sequence is shown in the right panel.*

More recently, it was stated that chromatin density acts as a barrier to the recruitment of DNA repair and DNA damage signalling proteins at sites of DNA damage (6). Therefore, specific histone modifications were previously suggested to induce recruitment of chromatin-interacting proteins involved in sensing DNA damage (7,8). A well-studied example is the phosphorylation of histone variant H2A.X by ATM kinase (9). Phosphorylated H2A.X acts as a beacon for proteins involved in the DNA damage response, like Tip60 (10) and MDC1 (11). The mechanism through which histone H2A.X is incorporated into nucleosomes and what triggers this process remains unclear.

Even though it is known that the histone NTTs provide a binding platform for a large number of proteins, the important mechanisms that regulate these interactions, and thereby coordinate chromatin remodelling events, are largely unknown. To address these issues we affinity-purified proteins interacting with core histone NTTs from cellular nuclear extracts. We studied the range of interactions between nuclear proteins and a core histone consensus NTT peptide, lacking any post-translational modifications. By using this model peptide that contains conserved amino acid motifs from core histone NTTs (H2A, H2B, H3 and H4, see Figure 1A and 1B), we anticipated to pull-down a broad set of histone NTT-interacting proteins. To investigate the dynamics of interactions, proteins were enriched by affinity pull-downs from nuclear lysates both of control cells and cells in which DNA double strand breaks were induced during 30 minutes. DNA DSBs were induced using bleomycin (12). Over 700 proteins were found to specifically bind to the immobilized peptide before and shortly after DNA damage induction. Next to the possible recruitment of specific proteins, the phosphorylation status of the histone NTT-interacting proteins was analysed using a two-fold strategy to enrich for phosphorylated peptides. This revealed that a number of proteins known to be involved in chromatin remodelling become either phosphorylated or dephosphorylated upon DNA damage induction. The identification of several new DNA damage-induced phosphorylation sites on histone chaperones suggests that DNA-related processes, such as nucleosome assembly, DNA replication, transcription and repair can be integrated (13) through the formation of chromatin remodelling protein complexes in the early onset of the DNA damage response, which is triggered by phosphorylation.

EXPERIMENTAL PROCEDURES

Cell cultures, DSB induction and preparation of nuclear extracts

Human lymphoblastoid cell culturing, methods for DNA damage induction and the purification of nuclei from human lymphoblastoid cells as well as the

subsequent extraction of proteins from nuclei were performed as described before (14). In brief, human lymphoblastoid cells were either not challenged or challenged using bleomycin for 30 minutes. After harvesting and washing, nuclei were purified using two sequential sucrose washes (2.4 M and 1 M, respectively). Nuclei were hypertonically lysed during 1 hour at 4°C and after centrifugation the supernatants were used for further study.

Peptide immobilization, enrichment and in solution digestion of histone-binding proteins

The core histone N-terminal tail consensus peptide SGRGKAGKKGRKGAKTRQC was immobilized via its C-terminal cysteine residue as described previously (14) (see Figure 1B). Beads were washed and stored in PBS containing 0.05% sodium azide. One milligram of nuclear protein lysate was buffer exchanged to IPH-E buffer containing 50 mM Tris.HCl, pH 8, 325 mM NaCl, 0,5% Triton X-100 and incubated with 25 μL of immobilized peptide-agarose slurry. After incubation on a rotating device for 1 hour at 4°C, the supernatant was removed and the beads were washed three times with IPH-E buffer. Bound proteins were subsequently dissolved in a solution of 8 M urea in 25 mM ammonium bicarbonate pH 8 and incubated with 750 ng endoproteinase LysC (Roche Diagnostics) for 4 hours at 37°C. Following reduction and alkylation using 2mM DTT and 4mM iodoacetamide respectively, the sample was diluted to 2 M urea with 50 mM ammonium bicarbonate pH 8 and incubated o/n with 750 ng trypsin at 37°C. In parallel, bound proteins were visualized using standard 1D SDS-PAGE.

Strong Cation Exchange Chromatography.

Strong cation exchange was performed using two Zorbax BioSCX-Series II columns (0.8 mm ID x 50 mm L, 3.5 μm), a Famos autosampler (LCpackings, Amsterdam, The Netherlands), a Shimadzu LC-9A binary pump and a SPD-6A UV-detector (Shimadzu, Tokyo, Japan). Prior to SCX chromatography protein digests were desalted using a small plug of C18 material (3M Empore C18 extraction disk) packed into a GELoader Tip similar to as previously described (15). The eluate was dried completely by vacuum centrifugation and subsequently reconstituted in 20% acetonitrile, 0.05% formic acid. After injection, the first 10 minutes were ran isocratically at 100% solvent A (20% acetonitrile, 0.05% formic acid, pH 3.0), followed by a linear gradient of 1.3% min^{-1} solvent B (500 mM KCl in 20% acetonitrile and 0.05% formic acid, pH 3.0). A total number of 25 SCX fractions (1 min each, *i.e.* 50 μL elution volume) were manually collected and dried in a vacuum centrifuge.

TiO_2 purification of phosphorylated peptides

The early eluting fractions from the SCX separation were reconstituted in 50% acetonitrile, 5% formic acid and subjected to phosphopeptide purification using TiO_2 (21). After loading the sample onto the TiO_2 microcolumn, the column was washed with 20 μL 50% acetonitrile and 5% formic acid. Phosphopeptides were eluted using 10 μL of 1.25% ammonia in water (pH 10.5) and directly mixed with 10 μL 2% formic acid in water. The flow-through of the TiO_2 microcolumns was dried to completeness, reconstituted in 0.1 M acetic acid in water and subjected to nanoLC-MS analysis.

On-line nanoflow liquid chromatography FT-ICR-MS.

Residues were reconstituted in 10 μL 0.1 M acetic acid and were analyzed by nanoflow liquid chromatography using an Agilent 1100 HPLC system (Agilent Technologies, Waldbronn, Germany) comprising of a solvent degasser, a binary pump, and a thermostated wellplate autosampler, coupled on-line to a 7-Tesla LTQ-FT mass spectrometer (Thermo Electron, Bremen, Germany). The liquid chromatography part of the system was operated in a set-up essentially as described previously (16). Aqua™ C18, 5 μm, (Phenomenex, Torrance, CA, USA) resin was used for the trap column and ReproSil-Pur C18-AQ, 3 μm, (Dr. Maisch GmbH, Ammerbuch, Germany) resin was used for the analytical column. Peptides were trapped at 5 μL /min in 100% solvent A (0.1 M acetic acid in water) on a 2 cm trap column (100 μm internal diameter, packed in-house) and eluted to a 25 cm analytical column (50 μm internal diameter, packed in-house) at ~150 nl/min in a 50-min gradient from 0 to 40% solvent B (80% acetonitrile, 0.1 M acetic acid). The eluent was sprayed via emitter tips (made in-house), butt-connected to the analytical column. The mass spectrometer was operated in data dependent mode, automatically switching between MS and MS/MS and neutral loss driven MS^3 acquisition. Full scan MS spectra (from m/z 300-1500) were acquired in the FT-ICR with a resolution of 100,000 at m/z 400 after accumulation to target value of 500,000. The three most intense ions at a threshold above 5000 were selected for collision-induced fragmentation in the linear ion trap at a normalized collision energy of 35% after accumulation to a target value of 15,000. The data dependent neutral loss settings were chosen to trigger a MS^3 event after a neutral loss of either 24.5, 32.6 of 49 ± 0.5 m/z units was detected amongst the 5 most intense fragment ions.

Data analysis

All MS^2 and MS^3 spectra from each LC run were merged to a single file which was searched using the Mascot search engine (Matrix Science) against the

SwissProt database (version 48.3) with carbamidomethyl cysteine as fixed modification, protein N-acetylation, oxidized methionines, and phosphorylation of serine, threonine or tyrosine as variable modifications. Trypsin was specified as the proteolytic enzyme and up to two missed cleavages were allowed. The mass tolerance of the precursor ion was set to 10 ppm and that of fragment ions was set to 0.5 Da. All phosphorylated peptides identified during Mascot searches were confirmed by manual interpretation of the spectra. Characteristics of all proteins in the set that have a minimal protein score (Mascot) of 60 and an individual peptide score of 20 (both scores are calculated by the database interface program and are based on the probability of a protein/peptide sequence identification, $p<0.05$) within an accuracy of 10 ppm. After exclusion of cytoskeletal, ribosomal and spliceosomal proteins, the analysis set was further focused by applying a semi quantitative ranking based on protein sequence coverage, for which the following criteria were used: at least 1 unique confident peptide identification *per* 100 amino acids in a protein (for proteins smaller than 100 kDa) or at least 5 confident peptide identifications (for proteins larger than 100 kDa) had to be detected. These criteria correct for the fact that larger proteins in general get higher protein scores.

RESULTS

Pull-down of core histone N-terminal tail binding proteins using affinity proteomics

Numerous proteins that bind to core histone N-terminal tails have been identified and their specific roles in for example nucleosome assembly and/or transcriptional regulation have been characterized (for a review see: (17)). To gain insight into dynamic chromatin-protein interactions at a more comprehensive scale, the entire subset of proteins interacting with core histone N-terminal tails (NTT) was studied simultaneously using an affinity proteomics approach. Therefore, a peptide was designed that resembled a consensus core histone N-terminal tail sequence (Figure 1A). The peptide is built up from the 5 H2A/H4 N-terminal residues (SGRGK), two dibasic (KK and RK) motifs, which occur in H2B and H3, the AK motif from H2A and H4 and finally the RQ motif, as present in the H3 NTT. Following peptide immobilization according to the method shown in Figure 1B, proteins that were affinity-purified from nuclear lysates of human lymphoblastoid cells were analysed using in solution two-step digestion (endoproteinase LysC and trypsin), multidimensional nanoLC and FT (tandem) mass spectrometry. We focused on the nuclear proteome because the

initial response to DNA damage is expected to occur in the nucleus and in this way contamination of aspecifically binding cytosolic proteins was reduced.

Analysis of proteins that bind to the immobilized N-terminal tail peptide.

To get an idea of the composition of the subproteome enriched via the histone N-terminal tail affinity purification, a 1D SDS-PAGE separation of the fractions obtained throughout the enrichment procedure was run, which is shown in Figure 2.

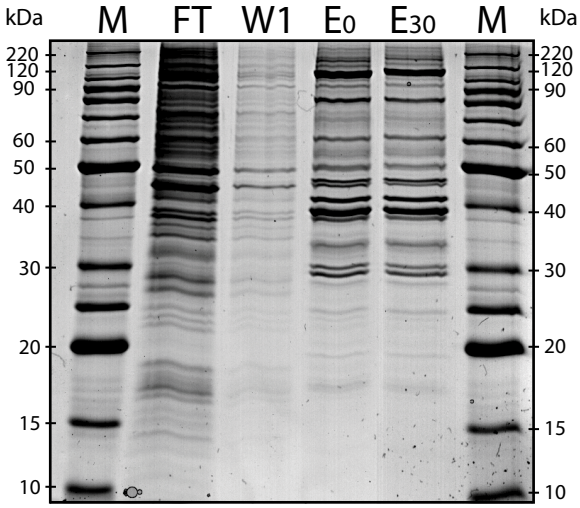


Figure 2. *1D SDS-PAGE separation of proteins enriched from a nuclear lysate of human lymphoblastoid cells (M, molecular weight marker; FT, flow through; W, wash fraction; E_0/E_{30}, eluate fractions enriched from control cells and cells in which DNA damage was induced during 30 minutes).*

From this image we conclude that our affinity purification enriches for specific proteins, as compared to the flow through fraction (*marked FT*). Even though LC-MS/MS analysis of the in gel digested proteins yielded a good overview of the subset of pulled down proteins (data not shown), a subsequent conducted (gel-free) two-step digestion analysis provided a better coverage of proteins binding to the NTT compared to the 1D-PAGE-MS analysis, especially with respect to high molecular weight proteins. The gel-free approach enabled identification of over 700 (compared to 530 in the gel-based analysis) proteins binding to the immobilized peptide. All identifications from the gel-free analysis are given in Supplementary Table 1. To further focus on proteins of interest, we implemented a semi-quantitative ranking, as described in the Experimental Section, which was

based on the requirement for a protein to be identified by at least one unique peptide per 100 amino acids (≈ 10 kDa). Cytoskeletal, ribosomal and spliceosomal proteins covered approximately 40% of the proteins enriched. Even though some of these proteins have been reported to associate with chromosomes (18) they will be excluded from the following discussion, as we were primarily interested in proteins more directly involved in chromatin remodelling. An overview of the composition of the final analysis set is given in Figure 3.

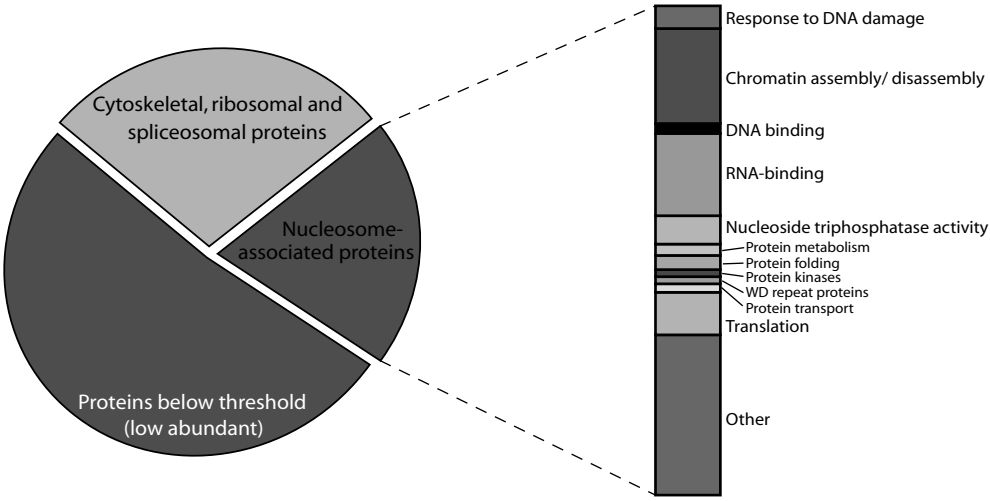


Figure 3. *Classification of the subset of proteins discussed in the text that was obtained after the semi-quantitative ranking and subsequent exclusion of cytoskeletal, ribosomal and spliceosomal proteins, as described in the Experimental Section.*

The qualitative differential binding of proteins to the histone NTT peptide was evaluated by comparing the set of binding proteins enriched from a nuclear lysate of control cells to that of a nuclear lysate of cells in which DNA DSBs were induced during 30 minutes. The subset enriched from the cells in which DNA DSBs were induced, consisted of 264 proteins while 260 proteins were detected in the sample enriched from the control cells. The two datasets shared 220 proteins; so 44 proteins were uniquely detected after DNA DSB induction, whereas 40 were only detected before DNA damage induction. An overview of the protein scores for the proteins identified in both runs is given in Figure 4. The plot shows that the majority of proteins is found in both samples with the same score ($y = 1.05x$, R^2 0.87, see Figure 4). It illustrates that the two individual analyses are highly similar and reproducible. The protein score was used as a

semi-quantitative indication of protein abundance, although we realize that this is only an assumption.

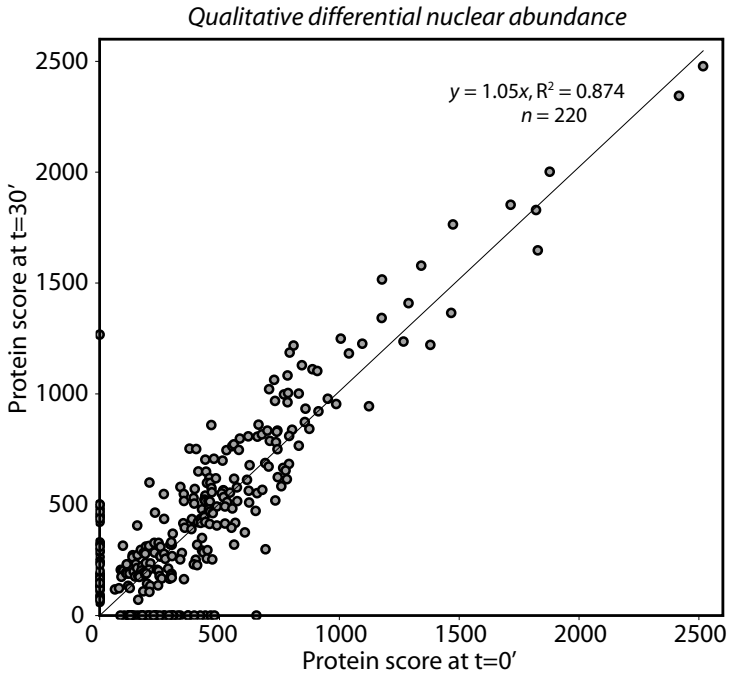


Figure 4. *Comparative proteome analysis of control cells (t=0') and cells in which DNA DSBs were induced using bleomycin during 30 minutes (t=30'). The plot of the scores of all proteins identified in the t=0' sample versus the t=30' sample reveals that some proteins are only found at one of the timepoints (represented by the dots that lie on the x- and y-axis, respectively). The majority of the proteins however is found in both sets (for examples see Table 1). Moreover, it illustrates the excellent reproducibility of the two individual analyses: the protein scores correlate well.*

Many known histone-/nucleosome-interacting proteins were found to be enriched, exemplified by the list of identified proteins given in Table 1, including all three subunits of a histone deacetylase complex (SAP18, HDAC1 and SIN3a) that enhances transcriptional repression, histone acetyltransferase type B catalytic subunit and subunit 2, nucleosome assembly proteins and chromatin assembly factor 1. Proteins that were apparently more abundant before DNA damage induction include Msx2-interacting protein, ubiquitin-protein ligase EDD1 and the ATP-dependent RNA helicase DDX3X. The PITSLRE serine/threonine protein kinase CDC2L2, Matrin-3 and hnRNP H' were only identified after DNA damage induction. However so far, none of these proteins have been linked to specific DNA damage-related processes.

Table 1. *Examples of proteins that were identified after enrichment from nuclear lysates of human lymphoblastoid cells using the immobilized histone N-terminal tail peptide. The code in parentheses is the SwissProt/TrEMBL accession number, followed by the corresponding protein name. The full set of identified proteins is given in Supplementary Table 1.*

Identified proteins enriched from cellular nuclear lysates using the immobilized consensus histone N-terminal tail peptide
(Q9UKV3) Apoptotic chromatin condensation inducer in the nucleus
(P39687) Acidic leucine-rich nuclear phosphoprotein 32 A
(P21127) PITSLRE serine/threonine-protein kinase CDC2L1
(O14646) Chromodomain-helicase-DNA-binding protein 1
(O14929) Histone acetyltransferase type B catalytic subunit
(O00422) Histone deacetylase complex subunit SAP18
(Q13547) Histone deacetylase 1
(Q96ST3) Paired amphipathic helix protein Sin3a
(P12956) ATP-dependent DNA helicase II, 70 kDa subunit
(P13010) ATP-dependent DNA helicase II, 80 kDa subunit
(P55209) Nucleosome assembly protein 1-like 1
(O75607) Nucleoplasmin-3
(P78527) DNA-dependent protein kinase catalytic subunit
(Q9NS91) Postreplication repair protein RAD18
(Q09028) Chromatin assembly factor 1 subunit C
(O60264) SWI/SNF related matrix associated actin dependent regulator of chromatin

Supplementary Table 1. *Overview of all proteins identified in the analysis run of the sample enriched from the lysate of control cells (t=0') and that of cells in which DNA damage was induced for 30 minutes (t=30'). 'Access'= accession number of the protein in the SwissProt database, 'Mw'= theoretical molecular weight of the protein, 'Score'= Mascot score of the protein, 'Pept'= number of peptides used for identification of the protein, 'Protein name'= protein name(s) and identifier from the SwissProt database.* **This table can be found on the supplementary CD-rom**

Post-translational modification of histone chaperones upon DNA DSB induction

As chromatin remodelling in the onset of the DNA damage response has to occur fast and transcription and translation are time-consuming processes, other activation mechanisms may regulate protein activity in the early response to DNA DSBs. Protein phosphorylation has been described as an activation mechanism for proteins in large numbers of pathways, including DNA repair (19). Therefore, the role of protein phosphorylation in remodelling events in the early response to DNA damage was studied. We enriched for phosphorylated peptides through SCX at low pH (20) followed by TiO_2 chromatography (21).

Table 2. *Phosphopeptides uniquely identified prior to, and after DNA damage induction. Entries marked with 'MA' reflect identifications that were manually annotated from MS/MS/MS spectra. Since not all of the identified phosphorylation sites are known, a kinase prediction was performed using the NetPhosK algorithm (http://www.cbs.dtu.dk/services/NetPhosK/). 'No pred' indicates that NetPhos gave no prediction for a particular phosphorylation site. CK1/2 = casein kinase 1/2 , PLK1 = Polo-like kinase, p38MAPK = p38 mitogen-activated protein kinase, PKG = cGMP dependent protein kinase, GSK3 = Glycogen synthase kinase-3.*

Phosphopeptides uniquely identified in the untreated cells (t=0')			
Protein ID Mass (Da)	**Score**	**Peptide**	**Consensus site prediction**
NUCL_HUMAN (P19338) Nucleolin			
1640.63948	57	K.GFGFVDFN**S^{618}**EEDAK.E	S^{618}: CK1 / CK2
YBOX1_HUMAN (P67809) Nuclease sensitive element binding protein 1			
1767.67366	39	R.NYQQNYQN**S^{165}**ESGEK.N	S^{165}: CK2
NPM_HUMAN (P06748) Nucleophosmin (NPM)			
2306.17447	44	K.MSVQPTV**S^{88}**LGGFEITPPVVLR.L	S^{88}:*No pred.*
3044.21350	52	-.MED**S^{4}**MDMDMSPLRPQNYLFGCELK.A + M_{Ox} ; Ac_{N-t}	S^{4}: CK2, PLK1
3028.21860	45	-.MEDSMDMDMS^{10}PLRPQNYLFGCELK.A + Ac_{N-t}	S^{10}: p38MAPK
NP1L4_HUMAN (Q99733) Nucleosome assembly protein 1-like 4 (NAP2)			
1939.78360	56	M.ADH**S^{4}**FSDGVPSDSVEAAK.N + Ac_{N-t}	S^{4}: *No pred.*

Phosphopeptides uniquely identified in cells in which DNA DSBs were induced (t=30')			
SSRP_HUMAN (Q08945) Structure-specific recognition protein 1			
2928.07050	64	K.EGMNPSYDEYAD**S^{444}**DEDQHDAYLER.M	S^{444}: CK2
CHD1_HUMAN (O14646) Chromodomain-helicase-DNA-binding protein 1			
1566.61984	106	R.RY**S^{1096}**GSDSDSISEGK.R	S^{1096}: PKG, GSK3
1646.58617	22	R.RY**S^{1096}**G**S^{1098}**DSDSISEGK.R	S^{1098}: *No pred.*
1549.16	*MA*	*R.RYSGSDSDSISEGK.R*	
SAFB1_HUMAN (Q15424) Scaffold attachment factor B			
1092.59433	23	K.SKGVPVIS**S^{576}**VK.T	S^{576}: CK1

Supplementary Table 2. *Overview of phosphorylated peptides identified from nuclear lysates of control cells (t=0') and that of cells in which DNA damage was induced for 30 minutes (t=30'). Phosphopeptides were identified both from MS/MS and MS/MS/MS fragmentation spectra. 'Observed'= peptide mass observed in the MS spectrum, 'Mr(expt)'= expected molecular weight based on peptide charge, 'Mr(calc)'= theoretical peptide mass, 'Delta'= difference between expected and calculated peptide mass, 'Score'= peptide score from Mascot, 'Expect'= expectance value assigned by Mascot, 'Peptide'= peptide sequence including the previous and next amino acid, to determine cleavage site. Proteins are represented as protein_ID, (accession number) protein name based on the SwissProt database.* **This table can be found on the supplementary CD-rom**

The use of a phosphorylation-specific mass spectrometric method (20) further aided the analysis of phosphorylated peptides. Prior to, and 30 minutes after DNA damage induction over 100 phosphopeptides were found, covering about 90 unique sites from 60 proteins (see Supplementary Table 2). The majority of the phosphorylated peptides was found in both nuclear extracts indicating that these sites are not significantly regulated in response to DNA DSB induction. More interesting, several 'unique' (induced after DNA damage induction) phosphopeptides were identified that may provide insight into the regulatory mechanisms of particular histone chaperones in the onset of the DNA damage response. A summary of 'unique' phosphopeptides is given in Table 2. As not all of phosphorylation sites given in Table 2 have been described before, a kinase consensus site prediction was performed to find possible kinase(s) for these sites, which are given in Table 2.

An example of DNA damage-induced phosphorylation is shown in Figure 5. This figure illustrates the relative change of a phosphorylated peptide from the structure-specific recognition protein 1 (SSRP1) compared to a co-eluting peptide of the Treacle protein (TCOF1) that did not change in abundance upon DNA DSB induction and this may be used as an internal standard. The fragmentation spectrum revealed that SSRP1 was phosphorylated on serine-444. This serine lies within a consensus sequence for casein kinase 2 (CK2, D**pS**D) embedded in the aspartate/glutamate-rich acidic domain of the protein.

In addition to serine-444 of SSRP1, two phosphorylation sites on CHD1 in a single tryptic peptide (serine-1096 and –1098), as well as serine-576 of scaffold attachment factor B1 (SAFB1) were found to be phosphorylated upon DNA DSB induction. Apparent DSB-induced phosphatase activity was detected as well: sites in the abundant nuclear phosphoproteins nucleolin and nucleophosmin, in addition to the acetylated N-terminal part of nucleosome assembly protein 1-like 4 (NAP-2) were dephosphorylated upon DNA DSB induction.

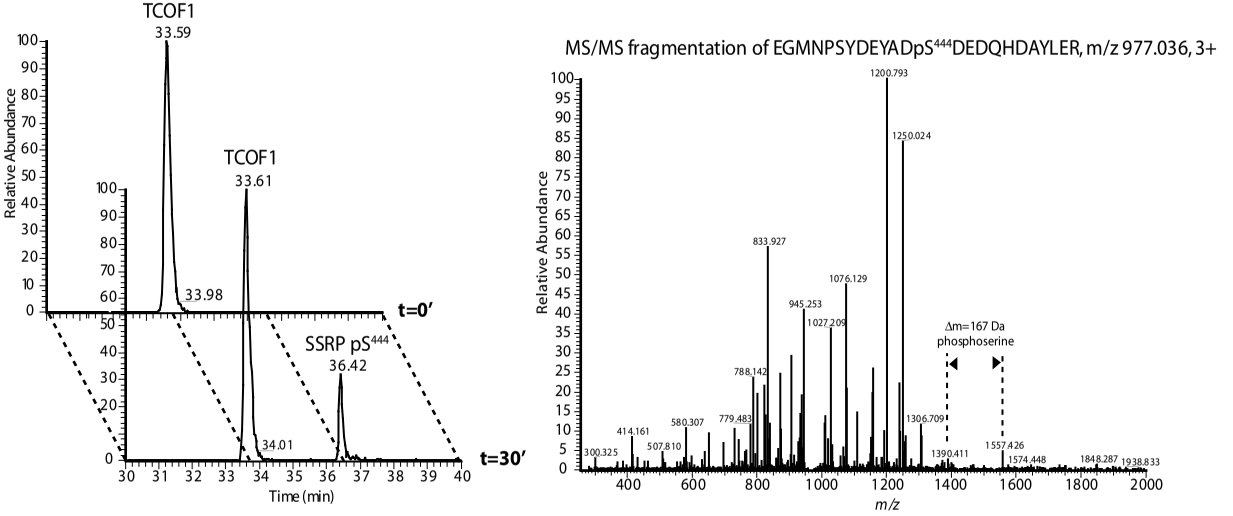


Figure 5. *Mass spectrometric identification of a phosphorylation site uniquely found upon DNA damage induction. In the left panel, the extracted ion chromatograms of the phosphorylated peptide from structure specific recognition protein 1 (SSRP1) together with that of a peptide originating from the Treacle protein (TCOF1) are given for both data sets. The abundance of the latter peptide did not change upon DNA damage induction and was therefore used as relative internal standard. This illustrates the regulation of the SSRP1 phosphorylation upon DNA damage induction (t=30'). On the right, the MS/MS fragmentation spectrum of the phosphorylated peptide EGMNPSYDEYADpSDEDQHDAYLER is given that revealed serine-444 to be the site of phosphorylation.*

DISCUSSION

Pull-down of core histone N-terminal tail binding proteins using affinity proteomics

Our study represents the first proteomic survey of the broad spectrum of interactions between proteins from a human nuclear lysate and the N-terminal tails (NTT) of core histones in nucleosomes. As can be seen both in Table 1 and Supplementary Table 1, many proteins that are known to interact with core histones were enriched through this approach thereby validating the set-up of our model system. Generally, domains in proteins that bind to histones within chromatin contain stretches of acidic amino acids (aspartic and glutamic acid) and those are present in many of the identified proteins. We also observed proteins that were most likely enriched via indirect association with the histone NTT peptide, since they are part of functional protein complexes of which only one constituent is known to interact with histone NTTs. This is illustrated by the identification of Ku80 that lacks an acidic histone-binding domain and was probably enriched indirectly through a physical interaction with Ku70. The dimer with Ku70, which contains two aspartate/glutamate-rich regions, is involved in the recognition of DNA lesions (22).

As illustrated by the 1D-PAGE image (shown in Figure 2) and the more comprehensive gel-free analysis, few qualitative differences were found in the protein sets enriched from control cells and cells in which DNA DSBs were induced. It is unlikely that these differences are a result of changes in protein expression within the time frame of the experiment (30 minutes), since transcription and translation are believed to take longer in higher eukaryotes. Therefore we expected that proteins uniquely identified either before or after DNA damage induction shuttle from, or are recruited to, the nucleus upon DNA damage induction. The marginal, though possibly interesting, contribution of shuttling to chromatin remodelling processes is not further discussed here.

Several proteins involved in chromatin remodelling are phosphorylated upon DNA damage induction

To enable a fast cellular response, the subset of proteins that binds to core histone NTTs on nucleosomes can be post-translationally modified, *e.g.* phosphorylated, in the early onset of the DNA damage response. This was found for several histone chaperones, such as the structure-specific recognition protein 1 on serine-444 (see Figure 5). The role of this identified phosphorylation site on SSRP1, which lies within a consensus sequence for casein kinase 2 (CK2, D**pS**D), is still unknown. Interestingly, other CK2-phosphorylation sites on SSRP1 have been related to UV-induced DNA damage induction, resulting in phosphorylation of

serine-510, -675 and -688 (23,24). The latter sites were stated *not* to be phosphorylated upon irradiation with γ-radiation, suggesting different regulatory mechanisms upon induction of different types of DNA damage (base oxidation and double strand DNA breaks resulting from UV- and γ-radiation, respectively). This is in line with our results, as these phosphorylation sites were indeed not detected here, which suggests that serine-444 phosphorylation is specific for the response to DNA DSBs.

In addition, two phosphorylation sites on the chromodomain-helicase-DNA-binding protein 1 (CHD1) were uniquely identified upon DNA DSB induction. Systematic mass spectrometric analysis, using MS^2 and MS^3, revealed that serine-1096 and serine-1098 within the same tryptic peptide, were the sites of phosphorylation. The peptide was found both singly and double phosphorylated. Not much has been reported about the effect of phosphorylation on the functioning of CHD1; therefore the potentially interesting role of phosphorylation on two sites upon DSB induction remains to be established. Scaffold attachment factor B1 (SAFB1) was found to become phosphorylated upon DNA damage induction on serine residue 576. This protein is responsible for the formation of a transcriptome complex by binding to S/MAR (scaffold/matrix attachment region) DNA.

The identified phosphorylation site lies within the part of the protein that both contains the nuclear localization signal and interacts with RNA polymerase II, indicating that this modification either regulates the cellular localization and/or the interaction with RNA polymerase II, thereby possibly repressing transcription in response to DNA damage induction. It is known that phosphorylation can greatly influence the activity of proteins and/or is a trigger for (the disruption of) specific protein-protein interactions. An example of such activation reported previously is the recruitment of phosphorylated chromatin assembly factor 1 (CAF1) to chromatin after UV irradiation of human cells, thereby linking chromatin assembly and DNA repair (25).

DNA damage-induced phosphatase activity detected on histone-interacting proteins

In addition to the DNA-damage-induced increase in phosphorylation, DNA damage specific decreases in phosphorylation were also detected: a number of phosphorylated peptides were only found in their phosphorylated forms in the lysates from control cells. An example of such a peptide is the acetylated N-terminal tryptic peptide (phosphorylated on serine-4) from NAP-2 that was found to be dephosphoryled upon DNA damage induction. Phosphorylation of this protein throughout the cell cycle was studied (26) and this showed that

dephosphorylation of NAP-2 triggers its transport into the nucleus. This however did not influence the binding of NAP-2 to histones, as was also found in this study. The protein was shown to be part of several multi-protein complexes together with for example histone H1 and CK2 or with DNA topoisomerase I (27), suggesting it is involved in several processes other than nucleosome assembly. NAP-2 is phosphorylated by CK2, but the phosphatase responsible for the dephosphorylation of this protein is not known.

Histone chaperones take part in multiple remodelling complexes enabling interplay between DNA-related processes.

In the affinity purification described here, primary interactions are expected to take place between stretches of acidic amino acids that occur in proteins and the basic amino acids present in the immobilized NTT peptide. This largely explains the binding of for instance SSRP1, SAFB1, NAP-2 and the β-subunit of CK2.

Other subunits of CK2 identified here, α and α', are probably enriched indirectly (whilst they are bound to subunit β). CK2 has not only been described to associate with nucleosomes during transcription (28) but was also reported to be involved in the DNA damage response by phosphorylating Mdm2 (29)

In our study, SSRP1 was found to be phosphorylated upon DNA damage induction on a serine that lies within a CK2 consensus sequence. SSRP1 has been reported to be a constituent of a multi-protein complex as is illustrated in Figure 6, containing CK2 that specifically phosphorylates serine-392 of p53 in response to DNA damage (30,31). Moreover as can be seen in Figure 6, SSRP1 has been described to be part of another complex, named FACT (*facilitates chromatin transcription*) that is involved in transcriptional regulation (32). This implicates a link between transcription and DNA repair via chromatin remodelling activities and is concomitant with previous reports on the involvement of histone chaperones in a variety of processes involving DNA, especially transcriptional regulation and DNA repair (33-35).

The enrichment of CHD1 through interaction with the NTT peptide is unexpected, since this protein does not contain acidic stretches. It does contain two chromodomains, which have been described to recognize methylated residues in histone NTTs (36). The fact that CHD1 was detected in both pull-down experiments indicates that the interaction is not influenced by the phosphorylation of serines 1096 and 1098. Nevertheless, this modification might recruit other proteins to this large histone chaperone and/or possibly triggers the formation of multi-protein complexes. The protein has not been related to the

DNA damage response, but it is clear that in transcriptional regulation CHD1 is preferentially located at active, decompacted regions in chromatin (37), which suggest that this protein helps to maintain chromatin in an open state. This is, of course, also highly favourable for the efficiency of DNA repair. Intriguingly, SSRP1 as well as SAFB1 have been described to interact with CHD1 (see Figure 6) and it is tempting to assume that these interactions are regulated by phosphorylation of CHD1. The interaction of CHD1 with SSRP1 occurs via an N-terminal segment of CHD1 that lies outside its chromodomain, but its exact function remains unknown (38). Yeast two-hybrid experiments conducted to reveal RNA polymerase II elongation factors yielded an interaction between CK2, SSRP1, hSPT16 and CHD1 (39). Finally, the interplay between SAFB1 and CHD1 was described to affect RNA splicing (40).

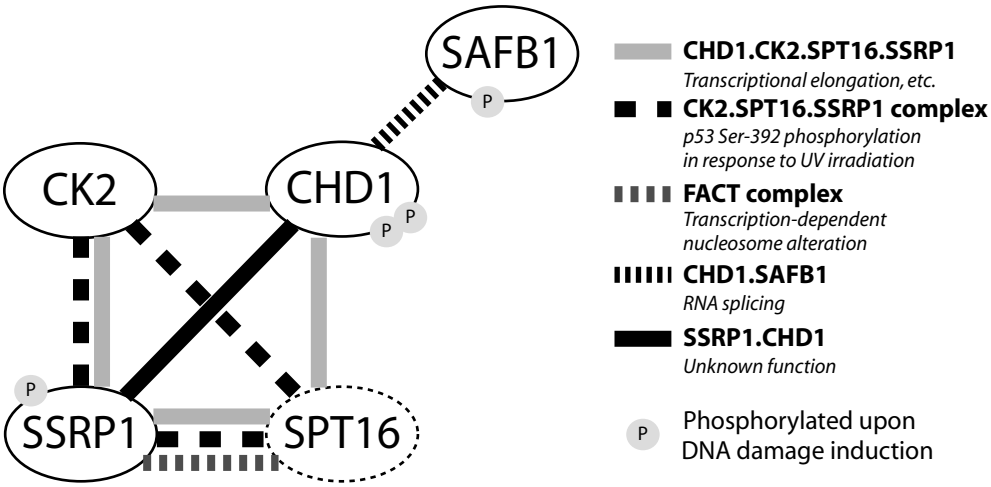


Figure 6. *Schematic representation of the interplay between histone chaperones that were differentially phosphorylated upon DNA damage induction. Several of the histone-binding proteins identified here, are known to be part of several protein complexes that act at several levels of chromatin remodelling and in various pathways, thereby coordinating DNA-related processes. This enables the integration of transcription, RNA splicing and DNA repair. Lines between proteins indicate that these are part of the same protein complex, and different types of lines resemble different protein complexes. CK2: casein kinase 2; CHD1: chromodomain-helicase-DNA-binding protein 1; SSRP1: structure-specific recognition protein 1; SAFB1: scaffold attachment factor B1, SPT16 (resembled by a circle with a dotted line because this protein was not found in our study): human ortholog of yeast suppressor of Ty insertion mutations.*

CONCLUSIONS

Taken together, the results of our affinity proteomics approach show that most histone chaperones involved in chromatin remodelling are constitutively present in the nucleus both prior to and after DNA damage induction. When remodelling is required, for example to allow DNA replication or transcription, specific histone chaperones are recruited to chromatin, which is triggered by post-translational modifications, such as phosphorylation. This was also shown for chaperones in the response to DNA damage: phosphorylation sites on SSRP1, CHD1 and SAFB1 were specifically found upon induction of DNA DSBs. These proteins have previously been described to be part of several protein complexes involved in multiple DNA-related processes of which an overview is given in Figure 6. Additionally, dephosphorylated sites were identified in histone chaperones, such as the nuclear phosphoproteins nucleolin and nucleophosmin as well as in NAP-2. The exact function of these modifications remains to be elucidated. We postulate that these post-translational modifications fulfil a role in the response to DNA DSBs at the level of chromatin remodelling preceding DNA repair, for example by disrupting or inducing protein-protein interactions on the core histone N-terminal tail platform.

Our approach provides an overview of histone N-terminal tail interacting proteins. Even though we find our approach to be highly sensitive in enrichment, it compromises specificity induced by specific histone sequences and/or post-translational modifications. Our data provides new insights into events that occur immediately upon DNA DSB induction featuring protein networks at the cross roads of nucleosome assembly, DNA replication, transcription and repair.

ACKNOWLEDGMENTS

This work was supported by the Netherlands Proteomics Centre.

REFERENCES

1. Morales, V., Giamarchi, C., Chailleux, C., Moro, F., Marsaud, V., Le Ricousse, S., and Richard-Foy, H. (2001) Chromatin structure and dynamics: functional implications. *Biochimie* **83**, 1029-1039
2. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-260
3. Berger, S. L. (2002) Histone modifications in transcriptional regulation. *Curr Opin Genet Dev* **12**, 142-148
4. Cheung, P., Allis, C. D., and Sassone-Corsi, P. (2000) Signaling to chromatin through histone modifications. *Cell* **103**, 263-271
5. Cosgrove, M. S., Boeke, J. D., and Wolberger, C. (2004) Regulated nucleosome mobility and the histone code. *Nat Struct Mol Biol* **11**, 1037-1043
6. Wuebbles, R. D., and Jones, P. L. (2004) DNA repair in a chromatin environment. *Cell Mol Life Sci* **61**, 2148-2153
7. Bilsland, E., and Downs, J. A. (2005) Tails of histones in DNA double-strand break repair. *Mutagenesis* **20**, 153-163
8. Hassa, P. O., and Hottiger, M. O. (2005) An epigenetic code for DNA damage repair pathways? *Biochem Cell Biol* **83**, 270-285
9. Fernandez-Capetillo, O., Lee, A., Nussenzweig, M., and Nussenzweig, A. (2004) H2AX: the histone guardian of the genome. *DNA Repair (Amst)* **3**, 959-967
10. Tyler, J. K. (2002) Chromatin assembly. Cooperation between histone chaperones and ATP-dependent nucleosome remodeling machines. *Eur J Biochem* **269**, 2268-2274
11. Stucki, M., Clapperton, J. A., Mohammad, D., Yaffe, M. B., Smerdon, S. J., and Jackson, S. P. (2005) MDC1 Directly Binds Phosphorylated Histone H2AX to Regulate Cellular Responses to DNA Double-Strand Breaks. *Cell* **123**, 1213-1226
12. Adema, A. D., Cloos, J., Verheijen, R. H., Braakhuis, B. J., and Bryant, P. E. (2003) Comparison of bleomycin and radiation in the G2 assay of chromatid breaks. *Int J Radiat Biol* **79**, 655-661
13. Chakravarthy, S., Park, Y. J., Chodaparambil, J., Edayathumangalam, R. S., and Luger, K. (2005) Structure and dynamic properties of nucleosome core particles. *FEBS Lett* **579**, 895-898
14. Dirksen, E. H., Cloos, J., Braakhuis, B. J., Brakenhoff, R. H., Heck, A. J., and Slijper, M. (2006) Human lymphoblastoid proteome analysis reveals a role for the inhibitor of acetyltransferases complex in DNA double-strand break response. *Cancer Res* **66**, 1473-1480
15. Gobom, J., Nordhoff, E., Mirgorodskaya, E., Ekman, R., and Roepstorff, P. (1999) Sample purification and preparation technique based on nano-scale reversed-phase columns for the sensitive analysis of complex peptide mixtures by matrix-assisted laser desorption/ionization mass spectrometry. *J Mass Spectrom* **34**, 105-116
16. Meiring, H. D., van der Heeft, E., ten Hove, G. J., and de Jong, A. P. J. M. (2002) Nanoscale LC-$MS^{(n)}$; technical design and applications to peptide and protein analysis. *J. Sep. Sci.* **25**, 557-568
17. Loyola, A., and Almouzni, G. (2004) Histone chaperones, a supporting role in the limelight. *Biochim Biophys Acta* **1677**, 3-11
18. Uchiyama, S., Kobayashi, S., Takata, H., Ishihara, T., Hori, N., Higashi, T., Hayashihara, K., Sone, T., Higo, D., Nirasawa, T., Takao, T., Matsunaga, S., and Fukui, K. (2005) Proteome analysis of human metaphase chromosomes. *J Biol Chem* **280**, 16994-17004
19. Wang, H., Guan, J., Perrault, A. R., Wang, Y., and Iliakis, G. (2001) Replication protein A2 phosphorylation after DNA damage by the coordinated action of ataxia telangiectasia-mutated and DNA-dependent protein kinase. *Cancer Res* **61**, 8554-8563
20. Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villen, J., Li, J., Cohn, M. A., Cantley, L. C., and Gygi, S. P. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A* **101**, 12130-12135

21. Pinkse, M. W., Uitto, P. M., Hilhorst, M. J., Ooms, B., and Heck, A. J. (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem* 76, 3935-3943
22. Jackson, S. P. (2002) Sensing and repairing DNA double-strand breaks. *Carcinogenesis* **23**, 687-696
23. Li, Y., Keller, D. M., Scott, J. D., and Lu, H. (2005) CK2 phosphorylates SSRP1 and inhibits its DNA-binding activity. *J Biol Chem* **280**, 11869-11875
24. Krohn, N. M., Stemmer, C., Fojan, P., Grimm, R., and Grasser, K. D. (2003) Protein kinase CK2 phosphorylates the high mobility group domain protein SSRP1, inducing the recognition of UV-damaged DNA. *J Biol Chem* **278**, 12710-12715
25. Martini, E., Roche, D. M., Marheineke, K., Verreault, A., and Almouzni, G. (1998) Recruitment of phosphorylated chromatin assembly factor 1 to chromatin after UV irradiation of human cells. *J Cell Biol* **143**, 563-575
26. Rodriguez, P., Pelletier, J., Price, G. B., and Zannis-Hadjopoulos, M. (2000) NAP-2: histone chaperone function and phosphorylation state through the cell cycle. *J Mol Biol* **298**, 225-238
27. Rodriguez, P., Ruiz, M. T., Price, G. B., and Zannis-Hadjopoulos, M. (2004) NAP-2 is part of multi-protein complexes in HeLa cells. *J Cell Biochem* **93**, 398-408
28. Guo, C., Davis, A. T., Yu, S., Tawfic, S., and Ahmed, K. (1999) Role of protein kinase CK2 in phosphorylation nucleosomal proteins in relation to transcriptional activity. *Mol Cell Biochem* **191**, 135-142
29. Allende-Vega, N., Dias, S., Milne, D., and Meek, D. (2005) Phosphorylation of the acidic domain of Mdm2 by protein kinase CK2. *Mol Cell Biochem* **274**, 85-90
30. Keller, D. M., Zeng, X., Wang, Y., Zhang, Q. H., Kapoor, M., Shu, H., Goodman, R., Lozano, G., Zhao, Y., and Lu, H. (2001) A DNA damage-induced p53 serine 392 kinase complex contains CK2, hSpt16, and SSRP1. *Mol Cell* **7**, 283-292
31. Keller, D. M., and Lu, H. (2002) p53 serine 392 phosphorylation increases after UV through induction of the assembly of the CK2.hSPT16.SSRP1 complex. *J Biol Chem* **277**, 50206-50213
32. Belotserkovskaya, R., Oh, S., Bondarenko, V. A., Orphanides, G., Studitsky, V. M., and Reinberg, D. (2003) FACT facilitates transcription-dependent nucleosome alteration. *Science* **301**, 1090-1093
33. Gontijo, A. M., Green, C. M., and Almouzni, G. (2003) Repairing DNA damage in chromatin. *Biochimie* **85**, 1133-1147
34. Frit, P., Kwon, K., Coin, F., Auriol, J., Dubaele, S., Salles, B., and Egly, J. M. (2002) Transcriptional activators stimulate DNA repair. *Mol Cell* 10, 1391-1401
35. Verger, A., and Crossley, M. (2004) Chromatin modifiers in transcription and DNA repair. *Cell Mol Life Sci* **61**, 2154-2162
36. Flanagan, J. F., Mi, L. Z., Chruszcz, M., Cymborowski, M., Clines, K. L., Kim, Y., Minor, W., Rastinejad, F., and Khorasanizadeh, S. (2005) Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature* **438**, 1181-1185
37. Lusser, A., Urwin, D. L., and Kadonaga, J. T. (2005) Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. *Nat Struct Mol Biol* **12**, 160-166
38. Kelley, D. E., Stokes, D. G., and Perry, R. P. (1999) CHD1 interacts with SSRP1 and depends on both its chromodomain and its ATPase/helicase-like domain for proper association with chromatin. *Chromosoma* **108**, 10-25
39. Krogan, N. J., Kim, M., Ahn, S. H., Zhong, G., Kobor, M. S., Cagney, G., Emili, A., Shilatifard, A., Buratowski, S., and Greenblatt, J. F. (2002) RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach. *Mol Cell Biol* **22**, 6979-6992
40. Tai, H. H., Geisterfer, M., Bell, J. C., Moniwa, M., Davie, J. R., Boucher, L., and McBurney, M. W. (2003) CHD1 associates with NCoR and histone deacetylase as well as with RNA splicing proteins. *Biochem Biophys Res Commun* **308**, 170-176

Summary

The genome, which holds all genetic information necessary for cellular functioning, is constantly challenged by exogenous and endogenous factors that can induce DNA damage. Multiple types of damage, such as strand cross-links and single strand DNA breaks, can threaten the viability of the cell and need to be detected and taken care of properly. Therefore, the cell is equipped with a number of mechanisms to detect and repair DNA damage and thereby prevent cell death. Even though the most severe type of damage, the DNA double strand break (DSB), also occurs during 'normal' cellular processes, like V(D)J recombination, it can have dramatic effects when remaining unrepaired. Although much is known about the role of proteins and protein complexes in the response to DNA DSBs, a comprehensive overview of all events occurring at the protein level upon DSB induction has never been reported. Proteomics, which is the analysis of the protein complement of the genome that is present in a cell at a certain moment in time, is ideally suited to bring this about. The aim of the work described in this thesis was to study the response of human lymphoblastoid cell lines to DNA DSB induction using different proteomics techniques. DSBs were induced using bleomycin, which is a radiomimetic compound that acts similarly to ionizing radiation and induces DNA double strand breaks.

Mapping of cellular events at the proteome level, such as the response to DNA double strand breaks, requires methods that allow accurate and reproducible quantitation of protein expression levels in complex biological systems. These should not only enable the study of large changes in expression levels, but also permit the precise determination of small changes both for high abundant proteins and proteins at low concentration that can be responsible for subtle, but potentially important, changes in protein activity. Currently, one of the most accurate methods to accomplish this in quantitative proteomics is through the incorporation of stable isotopes (such as ^{13}C or ^{15}N) into proteins during cell growth. This however, is not applicable to all types of cells and organisms and therefore alternative methods for protein quantitation have been developed. We compared such a relatively new, alternative 2D gel-based quantitation method: two-dimensional difference in-gel electrophoresis (2D-DiGE), to metabolic stable isotope labeling. For this direct comparison, discussed in **Chapter 2**, yeast (*Saccharomyces cerevisiae*) was grown in chemostat cultures under two nutrient limiting conditions, i.e. nitrogen- and carbon-limited, with either ^{14}N or ^{15}N as sole nitrogen source. This resulted in the incorporation of ^{14}N or ^{15}N nitrogen in all proteins, allowing mass spectrometry-based quantitation of protein expression levels. After extraction, proteins from the individual samples were labelled with fluorescent CyDyes, Cy3 or Cy5. The third dye available, Cy2, was used to label an internal standard consisting of equal amounts of both

samples. Following mixing of the doubly labeled samples, proteins were separated using 2D gel electrophoresis and protein expression levels were determined, using both a fluorescence scanner in combination with image analysis software and tryptic digestion combined with MALDI peptide mass fingerprinting. In-depth analysis of a characteristic set of proteins covering low and high molecular weight, acidic and basic pI's as well as protein isoforms, showed that both methods provide comparable results with excellent correlation when expression ratios are between -3.0 and 3.0. Outside this range, differences in background correction/signal-to-noise determination of the detection methods cause a deviation in quantitation of expression levels.

Subsequently as pointed out in **Chapter 3**, fluorescent labeling of proteins and 2D-DiGE was applied to the investigation of both the fast and more prolonged effects of bleomycin-induced DNA DSBs on the nuclear proteome of human lymphoblastoid cell lines. Firstly, this large-scale proteomic study emphasized the importance of accounting for genetic variation when working with human cell lines: inclusion of at least 9 human cell lines showed to provide consistent results in which experimental and biological variation could be distinguished. Secondly, it confirmed that labeling of protein samples with fluorescent CyDyes allows accurate quantitation of events in the response to DNA DSBs, as was illustrated by the regulation found for proteins known to be involved in the DNA damage response. Examples of the latter are Ku70 and HMG1 that are involved in the detection of DNA lesions and DNA repair, respectively. Interestingly, the nuclear levels of three proteins, which are known to form the inhibitor of acetyltransferase (INHAT) complex involved in transcriptional regulation through coordination of chromatin remodelling by binding to core histone N-terminal tails, were found to decrease rapidly upon DNA DSB induction. The interactions of these proteins with core histone N-terminal tails were further studied using an immobilized consensus peptide. This not only showed that INHAT proteins could be enriched using this approach, but also that binding of the proteins was strongly reduced upon acetylation of lysines in the consensus peptide. As it is known that acetylation of lysines (together with post-translational modifications on other amino acid residues) in core histone N-terminal tails opens up chromatin structure, allowing DNA repair enzymes access to the damaged DNA, our results suggest a role for the INHAT complex in chromatin remodelling events that precede DNA repair very similar to its role in transcriptional regulation that has been described by Seo, *et al* (1).

Of the lymphoblastoid cell lines derived from 14 individuals used in the study described, seven show a so-called hypersensitive phenotype. This means they are hypersensitive towards DNA damage induction, which is reflected by a

relatively high number of DNA double strand breaks per cell upon treatment with bleomycin. Since the underlying mechanisms for this decreased genome stability are unknown, we attempted to gain insight into the way these cells cope with DNA damage, which is described in **Chapter 4**. This revealed differences in nuclear protein levels between the normal and the hypersensitive cell lines before and after induction of DNA DSBs with bleomycin. Proteins of interest could be categorized either as significantly higher in the normal cell lines, significantly higher in the hypersensitive cell lines or showing a significant interaction between time (after DNA damage induction) and phenotype. Among these were high mobility group protein 2 (HMG2), of which homologues in yeast were recently described to be involved in genome protection (2), as well as the DNA damage sensor protein Ku80, the acidic leucine-rich nuclear phosphoprotein 32A, a constituent of the INHAT complex and Ras-family protein Rap1B. The latter was reported to be involved in squamous carcinomas, which, together with the regulation differences of other proteins, provides an interesting starting point for additional research on the mechanisms underlying the increased cellular sensitivity towards bleomycin-induced DNA damage in cells with a hypersensitive phenotype. Interestingly, some of the proteins that were found to be specific for the phenotypes are also involved in the DNA damage response in general, as reported in Chapter 3. A member of the INHAT complex, pp32A, for example, showed significant higher expression levels in cells with a hypersensitive phenotype. This suggests that differences in regulation of chromatin structure can influence cellular sensitivity towards DNA damage induction.

In the tight regulation of the fast response to DNA DSBs mentioned before, protein phosphorylation plays a crucial role. A number of kinases and phosphatases that take part in complex signal transduction pathways of the DNA damage response can put on, or take off, phosphate groups from proteins thereby greatly influencing protein activity. To study this, a method was developed for the specific enrichment of serine- and threonine-phosphorylated peptides from tryptic digests, which is described in **Chapter 5**. This method is based on the base-catalysed β-elimination of phosphate from a serine- or threonine-phosphorylated peptide, which results in the formation of a reactive α, β-unsaturated amino acid that is susceptible to nucleophilic attack. After optimisation of the elimination reaction using single amino acids and tripeptides, this reaction was exploited for the subsequent addition of a dithiol (ethanedithiol) and a custom made probe molecule consisting of four modules, to a synthetic phosphorylated octapeptide. The use of functionally different modules first of all allowed coupling of the probe molecule to the previously phosphorylated peptide. Secondly, it enabled specific enrichment of the labeled

peptide using the biotin-avidin affinity couple. Thirdly, to circumvent the negative influences of a biotin moiety on mass spectrometric analysis, an acid-labile linker was built into the probe that allowed recovery of the labelled peptide from the avidin. Finally, a linker in which stable isotopes, such as ^{2}H and ^{13}C could be inserted was included in the probe. We were able to enrich a phosphopeptide from a mixture of synthetic peptides that resembled a tryptic protein digest and on top of that, the site of phosphorylation could be retrieved from the fragmentation spectrum of the labelled peptide. Even though our method provided chemical specificity, it also uncovered some weaknesses of the approach, such as a rather low reaction efficiency leading to low sensitivity. Since the chemical approach chosen here requires multiple steps, optimisation of the individual reactions is of utmost importance. Therefore, alternatives are being developed for the individual constituents of the multi-component probe. A solid-phase purification strategy, for example, could be implemented (instead of a biotin moiety) by synthesizing the probe on a functionalised resin. In this way, problems associated with biotin-avidin based purification methods are circumvented. To meet requirements for accurate quantitation of phosphorylation evens, the linker part within the probe molecule described in Chapter 5 could be replaced by a leucine analogue. This allows labeling of the probe molecule with stable isotopes by incorporation of either a 'light' or a 'heavy' ($6*^{13}C$) leucine building block, which is more convenient compared to the reduction of certain structures with deuterated reagents to introduce ^{2}H atoms, as suggested previously. In addition to the larger, more suitable, mass difference, stable isotopes such as ^{13}C and ^{15}N also have advantages over ^{2}H, especially when considering liquid chromatography based analyses, since deuterated compounds tend to show different chromatographic behaviour, which is a problem in quantitative proteomics. Even though some of the alternatives suggested here have been partially implemented in phosphoproteomics studies (3,4), follow-ups on such studies have proven to be very rare, indicating that this apparently is no guarantee for success.

Finally in **Chapter 6** the study of DNA DSB-induced changes in the sub-proteome of histone-binding proteins is described that uses the immobilized core histone N-terminal tail consensus peptide introduced in Chapter 3. The importance of chromatin remodelling in the onset of DNA repair was already pointed out in Chapter 3. Despite the rapid increase in interest in the role of chromatin remodelling in the DNA damage response, not much is known about the exact mechanisms through which for example histone variants are deposited into nucleosomes near damage sites or how interaction between proteins and core histone N-terminal tails regulate chromatin dynamics. Our study of the

differential interactions between proteins and an unmodified core histone N-terminal tail consensus peptide showed that many of the histone chaperones (proteins that interact with histone N-terminal tails and thereby regulate chromatin remodelling) are mostly constitutively present in the nucleus and are only post-translationally modified, *e.g.* phosphorylated or dephosphorylated, when remodelling activities are required, like in response to DNA DSBs. This was shown for a number of proteins, such as structure-specific recogntion protein 1, chromodomain-helicase-DNA-binding protein 1 and scaffold attachment factor B1 that were found to be phosphorylated after DNA damage induction. Nucleophosmin and nucleosome assembly protein 1-like 4 are examples of proteins that were dephosphorylated after DNA damage induction. This gel-free multidimensional chromatography FT-ICR-MS^n approach yields reproducible quantitation results when using the Mascot score as a semi-quantitative measure. Our affinity proteomics approach was shown to be highly sensitive and to enrich for a clear subset of proteins. Still it compromises specificity when concerning the vast amount of post-translational modifications that has been reported to occur on histone N-terminal (and C-terminal) tails. Specific (combinations) of modifications also referred to as the 'histone code' can be a trigger for the formation or disruption of specific histone-protein interactions. It will be interesting to establish a library of specifically modified core histone N-terminal tail sequences could provide insight into the effects of certain modifications on the regulation of DNA-related events.

Some of the histone chaperones identified in our study have been described to take part in multiple remodelling complexes and the results of our affinity-based proteomics experiment focuses on some complex features of numeral protein networks at the cross roads of nucleosome assembly, DNA replication, transcription and repair.

Taken together, the work described in this thesis highlights the strength of (quantitative) proteomics when analysing complex cellular mechanisms in which numerous proteins take part in multiple pathways, such as the response to DNA double strand breaks. It shows that proper experimental design allows the study of complex phenomena to in human cell lines that are known to exhibit considerable intrinsic variation. It is known that cells respond to DNA damage by arresting the cell cycle and starting up DNA repair. We found that even earlier events are required to open up the condensed structure of chromatin in which DNA is stored in the nucleus. These chromatin remodelling events that are regulated by post-translational modifications, such as protein phosphorylation, enable damaged DNA to be repaired in chromatin that is condensed when stored in the nucleus. The proteomics studies conducted here emphasize the

importance of protein post-translational modifications and relocalization in the regulation of cellular events.

REFERENCES

1. Seo, S. B., McNamara, P., Heo, S., Turner, A., Lane, W. S., and Chakravarti, D. (2001) Regulation of histone acetylation and transcription by INHAT, a human cellular complex containing the set oncoprotein. *Cell* **104**, 119-130
2. Giavara, S., Kosmidou, E., Hande, M. P., Bianchi, M. E., Morgan, A., d'Adda di Fagagna, F., and Jackson, S. P. (2005) Yeast Nhp6A/B and mammalian Hmgb1 facilitate the maintenance of genome stability. *Curr Biol* **15**, 68-72
3. Chowdhury, S. M., Munske, G. R., Siems, W. F., and Bruce, J. E. (2005) A new maleimide-bound acid-cleavable solid-support reagent for profiling phosphorylation. *Rapid Commun Mass Spectrom* **19**, 899-909
4. Tseng, H. C., Ovaa, H., Wei, N. J., Ploegh, H., and Tsai, L. H. (2005) Phosphoproteomic analysis with a solid-phase capture-release-tag approach. *Chem Biol* **12**, 769-777

Samenvatting

Het menselijk lichaam is opgebouwd uit vele miljarden cellen. Deze cellen bevatten onder andere DNA waarop informatie staat opgeslagen die nodig is voor het correct functioneren van een cel. Al het DNA in een cel wordt ook wel het genoom genoemd en is in feite een aaneenschakeling van blauwdrukken voor duizenden eiwitten. Eiwitten zijn moleculen die alle processen in de cel, zoals bijvoorbeeld energieproductie en afbraak van afvalstoffen, uitvoeren. Afhankelijk van de informatie die van het DNA wordt afgelezen -en dus van de eiwitten die aangemaakt worden-, specialiseert een cel zich tot bijvoorbeeld een spier-, zenuw- of huidcel. De meeste cellen delen voortdurend om zo het weefsel waar ze deel vanuit maken te laten groeien en/of gezond te houden. Tijdens het delingsproces is het van vitaal belang dat de informatie die op het DNA opgeslagen ligt in zijn geheel en op de juiste manier wordt overgedragen op de dochtercel. De structuur van DNA wordt echter voortdurend aangetast door moleculen in cellen die beschadigingen kunnen veroorzaken. Deze moleculen kunnen enerzijds afkomstig zijn van de chemische reacties die tijdens allerhande processen in de cel plaatsvinden en anderzijds van buiten de cel binnendringen. Voorbeelden van de laatste categorie zijn: componenten uit tabaksrook of uit voeding, zoals acrylamide (in chips en friet) en alcohol. Deze stoffen kunnen een scala aan DNA-beschadigingen veroorzaken, zoals cross-links in DNA-ketens en enkelstrengs DNA-breuken, die het correct aflezen van DNA belemmeren en daarbij het voortbestaan van een cel kunnen bedreigen. Daarom moeten ze op tijd opgemerkt, en vervolgens op adequate wijze verholpen worden. Hiertoe is de cel uitgerust met een arsenaal aan mechanismen die in staat zijn DNA- schade te detecteren en te repareren, en zo celdood te voorkomen. Hoewel de meest ernstige vorm van DNA-schade, de dubbelstrengs breuk, ook noodzakelijk is voor bepaalde normale cellulaire processen zoals productie van antilichamen, kan het dramatische gevolgen hebben wanneer de controle erover verloren gaat. Ofschoon er veel bekend is over hoe cellen reageren op DNA dubbelstrengs breuken, is een totaaloverzicht van alle gevolgen die dit type schade op cellulair niveau kan veroorzaken nog nooit beschreven.

Eiwitten spelen ook een bepalende rol bij het detecteren en repareren van DNA-schade. Daarom kan de analyse van het complement van alle door het genoom geproduceerde eiwitten, ook wel het proteoom (**prote**ïne + gen**oom**) genoemd, inzicht verschaffen in deze gebeurtenissen. Het doel van het hier beschreven werk is dan ook om de respons van humane cellijnen op inductie van DNA dubbelstrengs breuken te bestuderen met behulp van verscheidene technieken binnen proteomics onderzoeken. Om DNA-schade te induceren is bleomycine, een chemische stof die dubbelstrengs breuken veroorzaakt, gebruikt.

Een belangrijke analytische techniek in proteomics onderzoek is massaspectrometrie. Hiermee kan het molecuulgewicht van allerhande moleculen, waaronder eiwitten, bepaald worden. Met behulp van deze massa-informatie kan de identiteit van een eiwit vastgesteld worden, wat een sleutel kan zijn tot een biologisch proces. Het in kaart brengen van cellulaire gebeurtenissen, zoals de reactie op dubbelstrengs breuken, vereist methoden die de nauwkeurig en reproduceerbaar de hoeveelheid aangemaakte eiwitten, het zogeheten eiwit expressieniveau, in een complex biologisch mengsel kunnen bepalen. Zo kan vastgesteld worden welke eiwitten meer, of juist minder, aanwezig zijn in cellen als gevolg van bepaalde biologische verandering, zoals DNA-schade-inductie. De methoden moeten niet alleen in staat zijn grote ('aan/uit') verschillen te meten, maar juist ook nauwkeurig kleine veranderingen in eiwitniveaus kunnen bepalen die vaak, hoe subtiel ook, belangrijke veranderingen in activiteit van eiwitnetwerken kunnen bewerkstelligen.

Eén van de meest nauwkeurige methoden om eiwit expressieniveaus te meten is het labelen van eiwitten met stabiele isotopen tijdens het groeien van cellen of organismen. Dit zijn 'zware' varianten van de atomen waaruit eiwitten normaal gesproken zijn opgebouwd. Zo kunnen 13-koolstof (^{13}C) en 15-stikstof (^{15}N) gebruikt worden in plaats van 12-koolstof en 14-stikstof. De isotopen beïnvloeden de structuur en de functie van het eiwit niet, maar kunnen wel gebruikt worden voor de kwantificering van eiwit expressieniveaus met behulp van massaspectrometrische technieken omdat door incorporatie van stabiele isotopen deze eiwitten zwaarder (1 massa eenheid per ingebouwd koolstof- of stikstofatoom) wegen dan de gewone eiwitten. Omdat het gebruik van 'zware atomen' zich niet leent voor gebruik in alle typen cellen en organismen zijn er alternatieve methodes voor de kwantificering van eiwit expressieniveaus ontwikkeld. Hoofdstuk 2 beschrijft de resultaten van het experiment waarin we de kwaliteiten van de gevestigde (stabiele isotoop-)methode vergeleken hebben met die van een relatief nieuwe, meer algemene methode die gebaseerd is op labeling van eiwitten met een (fluorescente) kleurstof: difference in-gel electrophoresis (2D-DiGE). Hiervoor is bakkersgist (*Saccharomyces cerevisiae*) gegroeid in chemostat culturen onder twee verschillende nutriënt-limiterende condities: zowel stikstof- als koolstoflimitering. Daarbij werd ofwel ^{14}N of ^{15}N als enige stikstofbron gebruikt, wat volledige incorporatie van deze isotopen in alle gist-eiwitten tot gevolg had en dus een direct vergelijk van beide condities mogelijk maakte. Na extractie van eiwitten uit de gistcellen werden ze ook nog gelabeld met de fluorescente labels (CyDyes), Cy3 of Cy5. Het derde fluorescente label, Cy2, is gebruikt om een interne standaard te labelen. Deze standaard bestaat uit gelijke hoeveelheden van beide monsters en kan gebruikt worden om

te corrigeren voor kleine experimentele verschillen waardoor de resultaten betrouwbaarder worden.

Na het mengen van de dubbel-gelabelde monsters, zijn de eiwitten in dit mengsel gescheiden met behulp van 2D gel electroforese zodat ze individueel bestudeerd kunnen worden. Vervolgens zijn de eiwitexpressieniveaus in beide monsters bepaald: enerzijds met behulp van een fluorescentie scanner en image analyse software (voor de fluorescente labels) en anderzijds met behulp van tryptische digestie en MALDI massaspectrometrie (voor de stabiele isotopen). Analyse van een karakteristieke set eiwitten met uiteenlopende eigenschappen toonde aan dat beide kwantificeringsmethoden vergelijkbare resultaten opleverden die prachtig correleren zo lang de expressie ratio's 3 keer toe- of afnamen. Daarbuiten liepen de resultaten van beide methoden uiteen omdat hier verschillen in het bepalen van de ruis, en dus ook de signaal/ruis-verhouding, optraden.

In Hoofdstuk 3 worden de resultaten beschreven van de toepassing van de fluorescente labeling van eiwitten en 2D-DiGE om zowel korte- als langeretermijn effecten te bestuderen van bleomycine-geïnduceerde DNA dubbelstrengs breuken op het proteoom in de kern van humane cellijnen. Deze, voor proteomics begrippen grootschalige, studie benadrukte enerzijds hoe belangrijk het is om rekening te houden met de genetische variatie die bestaat in humane cellijnen. Omdat mensen intrinsiek verschillen, verschillen hun cellen qua typen en hoeveelheden eiwit ook, onafhankelijk van DNA-schade-inductie. Het bleek dat wanneer minstens 9 cellijnen in de analyse werden betrokken, de resultaten consistent waren en dat experimentele en genetische variatie onderscheiden konden worden. Anderzijds bevestigden de gevonden reguleringen (=veranderingen in hoeveelheden eiwit) voor eiwitten waarvan bekend is dat ze een rol spelen in de DNA-schaderespons, zoals Ku70 en HMG1, dat fluorescente eiwit-labeling nauwkeurige kwantificering van veranderingen in eiwitexpressieniveaus als gevolg van dubbelstrengs DNA-breuken mogelijk maakte. Erg interessant is de gevonden regulering voor drie eiwitten die samen het INHAT complex vormen. Dit eiwitcomplex is betrokken bij het aflezen van DNA doordat het de DNA-structuur kan beïnvloeden via binding aan het N-terminale deel van histonen. Histonen zijn speciale eiwitten die een soort spoelen vormen waaromheen DNA gewonden zit, zodat het makkelijk kan worden opgeslagen in de cel. De hoeveelheid van de INHAT eiwitten nam na inductie van dubbelstrengs DNA-breuken snel af, wat een rol voor dit eiwit in de DNA-schaderespons suggereert. Om de interacties van deze eiwitten met het N-terminale deel van histonen verder te bestuderen werd een stuk van een histon N-terminus nagemaakt en op een vaste drager geïmmobiliseerd zodat hier

vervolgens mee gekeken kon worden welke eiwitten specifiek aan dit stuk eiwit binden. Dit experiment toonde niet alleen binding van de INHAT eiwitten aan dit peptide aan, maar liet ook zien dat deze interactie volledig verstoord wordt door post-translationele (chemische) modificaties op het peptide, zoals acetylering van lysine residuen. Het is bekend dat zo'n chemische verandering de structuur van het DNA kan openen. Dit verschaft enzymen die betrokken zijn bij het repareren van DNA-schade vervolgens toegang tot het beschadigde DNA. Onze resultaten suggereren dat het INHAT complex hierbij een rol speelt, zoals het dat ook doet bij het openen van de DNA-structuur voorafgaand aan transcriptie.

Van de 14 door ons in bovenstaande studie gebruikte cellijnen zijn er zeven afkomstig van mensen met een zogenoemd 'hypergevoelig' fenotype. Dat betekent dat de cellen van deze mensen veel gevoeliger zijn voor DNA-schade inductie dan normale cellen. Dit komt tot uiting in een relatief hoog aantal dubbelstrengs breuken in het DNA per cel na behandeling met bleomycine. In de praktijk betekent dit dat de personen van wie deze cellen afkomstig zijn een hogere kans hebben op het ontwikkelen van kanker in het algemeen. Dit neemt bovendien verder toe wanneer zij roken en/of alcohol drinken. Omdat de oorzaak voor deze hypergevoeligheid niet bekend is, hebben wij geprobeerd inzicht te krijgen in de mechanismen die ten grondslag liggen aan de reactie van beide typen cellen op DNA-schade. Hiervoor is de dataset, beschreven in Hoofdstuk 3, met behulp van andere statistische methoden, verder geanalyseerd. Dit zou verschillen in expressieniveaus van bepaalde eiwitten aan kunnen tonen tussen de hypergevoelige en de normale cellijnen voor en na inductie van dubbelstrengs DNA-breuken. Potentieel interessante eiwitten konden in drie categorieën ingedeeld worden: 1) eiwitten die significant hoger tot expressie komen in normale cellijnen, 2) eiwitten die significant hoger tot expressie komen in de hypergevoelige cellijnen of 3) eiwitten die een significante interactie vertonen tussen tijd (na bleomycine inductie) en fenotype. Een voorbeeld uit de eerste categorie is Ku80, een eiwit betrokken bij DNA-schadeherkenning. Acidic leucine-rich nuclear phosphoprotein 32A, dat deel uitmaakt van het INHAT complex en een eiwit uit de Ras-familie: Rap1B. zijn voorbeelden uit de tweede categorie. Van Rap1B is bekend dat het een rol speelt bij het ontstaan van plaveiselcel-carcinoom. High mobility group protein 2 (HMG2), waarvan voor homologen in gist onlangs werd gevonden dat ze betrokken zijn bij bescherming van het genoom laat een interactie tussen tijd en celtype zien en valt daarom in categorie 3. Samen met informatie over de andere significant verschillende eiwitten, vormen deze data een startpunt voor verder onderzoek naar de onderliggende mechanismen die hypergevoeligheid in de gebruikte cellijnen veroorzaken. Een interessant gegeven ligt in het feit dat enkele van de in

Hoofdstuk 4 beschreven eiwitten die in expressieniveau verschillen tussen normale en hypergevoelige cellijnen, ook gereguleerd werden na DNA-schade inductie met bleomycine, zoals beschreven in Hoofdstuk 3. Een eiwit uit het eerder genoemde INHAT complex, pp32A bijvoorbeeld, vertoonde significant hogere expressieniveaus in cellen met een hypergevoelig fenotype. Dit impliceert dat verschillen in de regulering van chromatine structuur een bijdrage zouden kunnen leveren aan de gevoeligheid voor DNA-schade van bepaalde individuen.

De eerder beschreven snelle veranderingen, onafhankelijk van verschillen in eiwitexpressieniveaus, die optreden in respons op dubbelstrengs DNA-breuken worden onder andere gereguleerd door kinases en fosfatases. Deze enzymen maken deel uit van complexe signaaltransductieroutes en zijn in staat aminozuren in eiwitten te voorzien, of te ontdoen, van een fosfaatgroep wat een groot effect op de activiteit van het betreffende eiwit kan hebben. Doordat meestal slechts een klein deel van de eiwitten een fosfaatgroep krijgt, is de analyse van deze eiwitten tussen alle andere eiwitten die geen fosfaatgroep dragen, meestal lastig. Daarom hebben we een methode ontwikkeld voor de specifieke verrijking van gefosforyleerde peptiden uit een complex mengsel. Deze methode, die gebruik maakt van de base-gekatalyseerde β-eliminatie van fosfaat (fosforzuur) uit gefosforyleerde peptiden, is beschreven in Hoofdstuk 5. Uit de eliminatie-reactie wordt een reactief α, β-onverzadigd aminozuur gevormd dat gemodificeerd kan via worden een zogeheten nucleofiele additiereactie. Na optimalisatie van de eliminatiereactieparameters met behulp van gefosforyleerde aminozuren en tripeptiden werd de reactie gebruikt om achtereenvolgens een dithiol (ethaandithiol) en een speciaal ontworpen probe molecuul aan een gefosforyleerd synthetisch octapeptide te koppelen. Dit probe molecuul dient als een vishaak waarmee voormalig gefosforyleerde peptiden specifiek opgezuiverd kunnen worden. Het biedt tevens de mogelijkheid tot het inbouwen van eerdergenoemde stabiele isotopen, zoals ^{2}H en ^{13}C, waarmee relatieve kwantificering van fosforylering mogelijk wordt. Met onze methode konden hoeveelheden van 1 nanomol synthetisch gefosforyleerd peptide specifiek verrijkt worden uit een mengsel van synthetische peptiden. Daarnaast kon met behulp van massaspectrometrische fragmentatie van het gelabelde peptide de precieze positie van de fosfaatgroep bepaald worden. Ondanks de chemische specificiteit van de ontwikkelde methode, werden enkele zwakke punten van een dergelijke benadering blootgelegd. Het lage rendement in de individuele reacties bijvoorbeeld beïnvloedt de gevoeligheid negatief waardoor een echte proteomics toepassing van de methode (nog) niet mogelijk is. Mogelijkerwijs is er met de implementatie van een aantal aanpassingen verbetering te halen. Hierbij kan bijvoorbeeld gedacht worden aan vervanging van de biotinegroep voor de

affiniteitszuivering van het gelabelde peptide door een vaste drager die eenvoudiger uit het reactiemengsel gewonnen kan worden. Dit verkleint mogelijk de kans op aspecifieke binding van peptiden aan biotine en beperkt het aantal stappen dat nodig is voor de verrijking. Daarnaast zou voor koppeling aan de sulfhydrylgroep op het voormalig gefosforyleerd peptide gekozen kunnen worden voor een iodoacetyl-gefunctionaliseerde leucine analoog. De inbouw van zowel de 'lichte' als de 'zware' ($^{13}C/^{15}N$-gelabeld) leucine bouwstenen maakt relatieve kwantificering van fosforyleringen mogelijk. Het hiermee geïntroduceerde massaverschil is niet alleen groter dan dat in de eerdergenoemde probe, maar levert waarschijnlijk minder problemen op tijdens de analyse met behulp van vloeistofchromatografie, wat een belangrijke basis is voor een goede kwantitatieve vergelijking. Enkele van deze aanpassingen zijn reeds toegepast in gelijksoortige chemische proteomics toepassingen, maar het feit dat vervolgstudies op dit onderwerp vrij zeldzaam zijn, geeft aan dat ze niet automatisch een sleutel zijn tot succes.

Tenslotte beschrijft Hoofdstuk 6 de analyse van door dubbelstrengs DNA-breuken-geïnduceerde veranderingen in het subproteoom van eiwitten die binden aan het N-terminale deel van histonen en daardoor de toegankelijkheid van DNA kunnen bepalen. Deze groep van eiwitten is bestudeerd met behulp van het in Hoofdstuk 3 beschreven peptide. In Hoofdstuk 3 werd reeds melding gemaakt van het belang van structuurveranderingen in chromatine na inductie van dubbelstrengs DNA-breuken. Hoewel hier toenemende interesse voor is, is er nog maar weinig bekend over de meer precieze mechanismen die deze hermodelleringsactiviteiten reguleren. Zo is er steeds meer interesse in hoe interacties tussen specifieke eiwitten en het N-terminale deel van histonen in nucleosomen de structuur van chromatine kunnen coördineren. Met het experiment beschreven in Hoofdstuk 6 hebben we getracht hier een verbeterd inzicht in te krijgen. De analyse van differentiële interacties tussen eiwitten en het geïmmobiliseerde peptide (als gevolg van DNA-schade) toont aan dat veel van de bij chromatine hermodellering betrokken histon chaperones (dat zijn eiwitten die interacties aangaan met het N-terminale deel van histonen en zo de structuur van chromatine kunnen beïnvloeden) altijd in de kern aanwezig zijn en dat deze, indien nodig, post-translationeel gemodificeerd (bijvoorbeeld gefosforyleeerd of gedefosforyleerd) worden. Dit is het geval wanneer hermodellering van chromatine om reparatie van DNA mogelijk te maken, vereist is, zoals werd aangetoond voor structure-specific recognition protein 1, chromodomain-helicase-DNA-binding protein 1 en scaffold attachment factor B1. Deze eiwitten werden specifiek gefosforyleerd na inductie van dubbelstrengs DNA-breuken. Nucleophosmin en nucleosome assembly protein 1-like 4 zijn twee voorbeelden

van eiwitten die gedefosforyleerd werden na DNA-schade-inductie. Van enkele van de gevonden histon chaperones is beschreven dat ze deel uitmaken van verschillende eiwitcomplexen en daarbij interacties aangaan met andere eiwitten. De resultaten van dit experiment stellen ons in staat de complexe eigenschappen van meerdere eiwitnetwerken die verschillende DNA-gerelateerde processen, zoals DNA-replicatie, transcriptie en reparatie met elkaar in verband brengen, te bestuderen.

Kortom, het in dit proefschrift beschreven werk laat de toegevoegde waarde van proteomics onderzoek zien in de analyse van complexe cellulaire mechanismen waarin grote hoeveelheden verschillende eiwitten functioneren. De cellulaire respons op DNA-schade is een voorbeeld van een dergelijk mechanisme en we hebben laten zien dat dit soort studies met behulp van de juiste experimentele setup zelfs in humane cellijnen, waarvan bekend is dat ze intrinsiek al veel verschillen vertonen, uitgevoerd kunnen worden. Daaruit is gebleken dat vóór aanvang van processen als cel cyclus regulatie en herstel van DNA-schade, nog eerdere acties vereist zijn, die ervoor zorgen dat beschadigd DNA gerepareerd kan worden zonder dat de gecondenseerde structuur van chromatine, waarin DNA wordt opgeslagen, daarbij tot last is. Tenslotte benadrukken de hier uitgevoerde proteomics experimenten nogmaals de invloed van post-translationele modificaties en relocalisatie van eiwitten in de regulering van cellulaire processen.

Curriculum vitae

Eef Dirksen werd geboren op 17 juli 1978 te Eindhoven. Na het behalen van het VWO diploma aan het van Maerlantlyceum te Eindhoven in 1996, werd begonnen aan de studie levensmiddelentechnologie aan de Landbouwuniversiteit Wageningen. Tijdens deze studie werden afstudeervakken gevolgd bij de vakgroep Levensmiddelenchemie (onder begeleiding van dr G.A. van Koningsveld) en bij de vakgroep Biochemie (onder begeleiding van dr G. Oudgenoeg). Tevens werd een onderzoeksstage gelopen bij DSM Food Specialties in Delft (onder begeleiding van dr C. Heeremans). In september 2001 werd het doctoraalexamen behaald. Vervolgens begon hij in februari 2002 als assistent in opleiding bij de sectie Biomoleculaire Massaspectrometrie van de Faculteiten Farmaceutische Wetenschappen en Scheikunde aan de Universiteit van Utrecht. In deze functie werd het in dit proefschrift beschreven onderzoek verricht onder leiding van dr. Monique Slijper en prof. dr. Albert J.R. Heck. Behaalde onderzoeksresultaten werden onder meer gepresenteerd tijdens een lezing op de lustrumbijeenkomst van de *Nederlandse Vereniging voor Massaspectrometrie* in Kerkrade (2004) en in de vorm van posters gepresenteerd op het Proteomic Forum in München (2003), de 12^{e} editie van de *International Conference on Second Messengers and Phosphoproteins* in Montreal (2004) en tijdens de jaarlijkse bijeenkomst van de *American Society for Mass Spectrometry* in San Antonio (2005). In februari 2006 is Eef begonnen als post-doc binnen het Netherlands Proteomics Centre.

List of publications

Eef H.C. Dirksen, Jacqueline Cloos, Boudewijn J.M. Braakhuis, Ruud H Brakenhoff, Albert J.R. Heck, Monique Slijper. Human lymphoblastoid proteome analysis reveals a role for the INHAT complex in DNA double-strand breaks response. *Cancer Research* **2006** 66 (3): 1473-80.

Pieter van der Veken*, Eef H.C. Dirksen*, Eelco Ruijter*, Dirk T.S. Rijkers, Rob M.J. Liskamp, Monique Slijper, and Albert J.R. Heck. Development of multifunctional probes for quantitative profiling of phosphorylated serine and threonine residues in peptides and proteins. *Chembiochem* **2005** 6 (12): 2271-80. *These authors contributed equally to this work.

Annemieke Kolkman*, Eef H.C. Dirksen*, Monique Slijper and Albert J.R. Heck, Double Standards in Quantitative Proteomics: Direct comparative assessment of Difference in Gel Electrophoresis (DiGE) and metabolic stable isotope labelling. *Molecular and Cellular Proteomics* **2005** 4: 255-66. *These authors contributed equally to this work.

Eef H.C. Dirksen, Martijn W.H. Pinkse, Dirk T.S. Rijkers, Jacqueline Cloos, Rob M.J. Liskamp, Monique Slijper, Albert J.R. Heck. Investigating the dynamic nature of the interactions between nuclear proteins and histones upon DNA damage using an immobilized peptide chemical proteomics approach. *Journal of Proteome Research* **2006**, in press.

Gideon Oudgenoeg, Eef H. C. Dirksen, Steen Ingemann, Riet Hilhorst, Harry Gruppen, Carmen G. Boeriu, Sander R. Piersma, Willem J. van Berkel, Colja Laane, Alphons G. Voragen, Horseradish peroxidase-catalyzed oligomerization of ferulic acid on a template of a tyrosine-containing tripeptide. *Journal of Biological Chemistry* **2002** 277 (24): 21332-40.

Dankwoord

Omdat proteomics onderzoek zich afspeelt op het grensvlak van onder andere biochemie, biologie en analytische chemie kan het alleen goed uitgevoerd worden als 'vraag en aanbod' van de verschillende disciplines op elkaar aansluiten en als de juiste mensen bij elkaar gebracht worden. Dit is om mij heen gebeurd en daarom heb ik de afgelopen vier jaar met veel plezier aan mijn onderzoek kunnen werken. Hiervoor wil ik graag een aantal mensen bedanken.

Allereerst mijn promotor Albert die, stevig aan het roer van de groep én het NPC, alles prima coördineert. Albert, de combinatie van het feit dat jij iets verder van mijn onderzoek afstond en je onuitputtelijke stroom ideeën leverde vaak nuttige en inspirerende discussies op. Bedankt voor je snelle correctiewerk en alle geboden mogelijkheden, inclusief die in mijn huidige post-doc baan binnen het NPC.

Monique, als mijn co-promotor en dagelijkse begeleidster zette jij de samenwerking met het VU Medisch Centrum op die een belangrijke basis voor dit onderzoek is geweest. Bedankt daarvoor en voor de vakkundige manier waarop je me in het begin van het project bekend maakte met proteomics: ik vergeet nooit meer hoe je een 2D gel moet runnen!

Zoals ik al zei is de samenwerking met het VUmc erg belangrijk geweest. Jacqueline, bedankt voor je enorme bleo-ervaring en de goede zorg voor de cellen. Ruud en Boudewijn: bedankt voor jullie input, onder andere tijdens de inspirerende discussies die mij niet alleen celbiologische, maar ook veel nieuwe statistische inzichten hebben opgeleverd.

In Utrecht heb ik met veel plezier samengewerkt met een aantal chemisch Legoërs van Medicinal Chemistry die met heel basale bouwstenen de mooiste dingen kunnen maken. Achtereenvolgens hebben Ronald, Pieter, en de Chemical Brothers Eelco en Alex geprobeerd de labeling van fosfopeptiden te realiseren. Bedankt voor jullie aanhoudende inzet, helaas is gebleken dat van de wegen die leiden naar ontrafeling van het fosfoproteoom, de chemische niet de gemakkelijkste is… Dirk, met wie ik gedurende mijn hele project heb samengewerkt: bedankt voor ál je chemische bijdragen en voor de discussies in de trein terug waarin niet alleen chemie, maar ook bijvoorbeeld de laatst gelezen boeken de revue passeerden.

Veel dank gaat uit naar de mensen die je dagelijks op de werkvloer omringen en waarmee je de welbekende ups en downs van een AIO-onderzoek deelt. Vanwege de overlappende tijdspanne van vier jaar en het verloop, ontmoet je, en werk je samen met, een boel mensen. Ik wil iedereen daarvoor hartelijk bedanken, een

aantal mensen in het bijzonder. Om te beginnen mijn paranimfen Annemieke en Martijn. Kolk, ik vond het erg leuk om jouw paranimf te zijn en heb van die dag erg genoten. Hopelijk kan jij straks hetzelfde zeggen! Bedankt voor de altijd gezellige samenwerking die onder meer leidde tot de oprichting van de bierclub, ontginning van het geheime plekje en natuurlijk het meten-met-twee-maten-verhaal (ik baal nog steeds dat het muntje op het moment suprème jouw kant op viel!). Martijn, bedankt voor de erg leerzame, maar zeker ook gezellige, samenwerking. Ik heb dankbaar gebruik gemaakt van je kennis van allerhande zaken en ben nog altijd onder de indruk van je doorzettingsvermogen en je muiskliksnelheid. Als voetballiefhebbers (AZ en PSV gaan prima samen) hadden we gelukkig wat afleiding als de resultaten even nét niet zo waren als gedacht. Ik hoop dat we ook na je 'transfer' contact houden, succes in Delft!

Michael en Esther, AIO-collega's en kamergenoten van het eerste uur: bedankt voor de gezelligheid op Z301. Es, jij bent ook bijna klaar: wat zal ik ontspannen toekijken de 18e… Michael, de laatste tijd heb ik je nauwelijks meer gezien omdat je bij Biologie bent gaan wonen (waar is de tijd dat we nog in één bed sliepen ;)?). Ik hoop dat de afronding van je onderzoek voorspoedig verloopt én dat de derde Eman-telg gezond ter wereld komt! Juud, ex-kamergenote: bedankt voor je gezelschap, óf wel in de trein óf in de auto als je me weer eens bijna tot aan de voordeur thuisbracht vanuit Utrecht.

Simone, fosfoproteomics-collega: bedankt voor de gezelligheid, onder andere tijdens de dooie uurtjes in het Employment Center van ASMS, waar voor ons al nauwelijks iets te werken viel. Ik weet zeker dat je WB-, IP- en chip-data binnenkort als een prachtige Danio-puzzel in elkaar gaan vallen, hou vol! Arjen, bedankt voor de zuivering van die enorme hoeveelheid PKG, wie weet komt ie binnenkort nog van pas… Ken je trouwens die mop van die collega's die een midweekje gingen skiën? Marco, bedankt voor je fotografische handigheid en je bereidheid om zo vroeg naar Utrecht CS te komen!

Mijn studenten, Sjoerd en Michiel, wil ik bedanken voor hun bijdrage aan mijn onderzoek. Ik vind het leuk dat jullie allebei je plek hebben gevonden!

Verder zeg ik tegen Anca, Andreas, Arijan, Bas, Bas, Cees, Clive, Corine, Dennis, Edwin, Harm-Jan, Hortense, Hugo, Jantien, Jeffrey, Jeroen, Jeroen, Jeroen, Johan, Joost, Kristina, Linda, (tjesus) Leon, Manuel, Martina, Mirjam, Nasrin, Natasja, Onno, Paul (Seal), Poupak, Renske (van de Blue Lagoon), Robert, Ron, Ronald, Shabaz, Sharon, Sylvia, Sylvie, Thin-Thin, Tieneke, Werner (Uh, weet jij waar Eef Dirksen zit?) en Wilma: bedankt, thanks, merci, Danke!

Uiteraard zijn er ook naast het werk een heleboel mensen die door hun interesse of medeleven, of beiden, belangrijk zijn geweest tijdens mijn vier AIO-jaren. De Eindhovense groep vrienden, die al vanaf de brugklas dateert en sinds die tijd alleen maar verder is uitgedijd, wil ik graag bedanken voor al hun afleiding in de vorm van, bijvoorbeeld, vrijdagavond muntjes drinken in de Thomas, weekendjes Ardennen (waar ik al weken vantevoren naar uit kijk), zomervakanties (Argelès zal ik niet snel vergeten!), voetbal kijken en zo kan ik nog wel even doorgaan. Hopelijk gaat dit nog heel lang zo verder!

Studiegenoten Miriam, Arjon en René: ik vind het leuk dat ons contact, ondanks de drukke agenda's, blijft bestaan. Arjon (en Myrthe), bedankt voor de gastvrijheid tijdens de keren dat ik in Utrecht bleef slapen. Ik hoop dat er aankomend seizoen in vak GG van het Philips Stadion weer flink te genieten valt! Ook de traditie van gastronomische topavonden moet zeker in stand gehouden worden. Gideon, hoewel je meer Amsterdammer (of Londonner) bent, val je ook onder het kopje Wageningen. Grappig dat mijn eerste MS-ervaringen op de LCQ tijdens mijn afstudeervak bij jou werden geboren uit problemen met de NMR. Ik heb veel van je geleerd, prettig met je samengewerkt (kijk maar naar de sleeping beauty) en ik waardeer de tijd die we samen doorbrengen altijd zeer. Leuk dat je voor mijn promotie overkomt!

De mannen van het zaterdagochtendvoetbal: bedankt voor de voetbaltechnische (en fysieke) weerstand die jullie bieden, het is altijd weer heerlijk om anderhalf uur in de wei te draven.

Janny, Toon, Pieter en Anne: bedankt voor de interesse en gezelligheid!

Lieve papa en mama, Alexia en Rob, Paul en Anja. Jullie zijn natuurlijk al veel langer dan de afgelopen vier jaar heel belangrijk voor me. Ontzettend bedankt voor jullie steun, interesse, het eindeloze vertrouwen in mijn (wetenschappelijke) kunnen en (waar nodig) opbeurende woorden gedurende mijn AIO-tijd. Het is een prettig gevoel te weten dat jullie er altijd voor mij zijn!

Lieve Quirine en Philine, dankzij jullie schattigheid en totale onbekendheid met de fenomenen 'AIO' of 'promoveren' vormden jullie altijd de perfecte afleiding!

Liefste Marchien, volgens mijn belofte moet dan nu ons echte leven weer beginnen! Ontzettend bedankt voor al je steun en begrip tijdens de momenten dat ik even minder prettig in de omgang was (wat je altijd perfect aanvoelt) en voor al je andere heerlijke eigenschappen, zoals je zorgzaamheid en sociaal gevoel. Ik ben erg trots op je en ik hoop dat je recente carrière-move zo uit gaat pakken als je je voorstelt want je verdient het. Luv joe!