

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

**Phosphoproteomics and proteomic phenotyping to
assess signal transduction in cancer cells**

Cuiping Pan

aus

Guang Dong, VR China

2008

Erklärung

Diese Dissertation wurde im Sinne von §13 Abs. 3 der Promotionsordnung vom 29. Januar 1998 von Herrn Prof. Matthias Mann betreut.

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig, ohne unerlaubte Hilfe erarbeitet.

München, am 30.06.2008

Cuiping Pan

Dissertation eingereicht am 02.07.2008

1. Gutachter: Prof. Dr. Matthias Mann
2. Gutachter: Prof. Dr. Patrick Cramer

Mündliche Prüfung am 30.07.2008

Table of Contents

| | |
|--|-----|
| Abbreviations | III |
| Summary | 1 |
| 1 Introduction..... | 3 |
| 1.1 Signal transduction in cancer | 3 |
| 1.1.1 Signal transduction paradigm | 3 |
| 1.1.2 Mechanism of signal transduction | 4 |
| 1.1.3 Altered signal transduction in cancer..... | 6 |
| 1.1.4 Kinases, phosphatases and cancer | 8 |
| 1.1.5 Molecular targeted cancer therapies | 12 |
| 1.2 Mass spectrometry in biology | 16 |
| 1.2.1 General workflow of proteomics | 16 |
| 1.2.2 Mass spectrometric instrumentation | 18 |
| 1.2.3 Quantitative proteomics | 30 |
| 1.2.4 Phosphoproteomics | 33 |
| 2 Aim of the Study..... | 37 |
| 2.1 Cancer cell line to evaluate inhibitors for kinases and phosphatases | 37 |
| 2.2 Liver cell model to assess normal and transformed cells | 37 |
| 3 Materials and Methods..... | 38 |
| 3.1 Cell models | 38 |
| 3.1.1 Immortalized cell lines..... | 38 |
| 3.1.2 Primary hepatocytes..... | 39 |
| 3.1.3 Cell culture..... | 41 |
| 3.1.4 Preparation of inhibitor compounds | 42 |
| 3.1.5. Cell stimulation, harvest and protein recovery | 42 |
| 3.2 Proteome mapping | 44 |
| 3.2.1 Sample preparation | 44 |
| 3.2.2 Mass spectrometric methods..... | 46 |
| 3.3 Phosphoproteome mapping | 47 |
| 3.3.1 Sample preparation | 47 |
| 3.3.2 Mass spectrometric methods..... | 49 |
| 3.4 Mass spectrometric data analysis..... | 50 |

Table of Contents

| | | |
|-------|--|-----|
| 3.4.1 | Identification | 50 |
| 3.4.2 | Quantitation | 51 |
| 3.4.3 | Assign localization of the phosphate group | 52 |
| 3.4.4 | Phosphoproteomic dataset stored in PHOSIDA database | 52 |
| 3.5 | Bioinformatic analysis..... | 52 |
| 3.5.1 | Gene Ontology and KEGG enrichment analysis | 52 |
| 3.5.2 | Phosphorylation - motif check | 53 |
| 3.5.3 | Phosphorylation - structural analysis | 54 |
| 3.5.4 | Phosphorylation - phosphorylation site predictor | 54 |
| 4 | Results and Discussions | 55 |
| 4.1 | Project 1 - Quantitative phosphoproteome analysis of a liver cell line reveals specificity of phosphatase inhibitors | 55 |
| 4.1.1 | Evaluation of the phosphoproteomics technique | 58 |
| 4.1.2 | Phosphoproteome of the mouse liver cell line Hepa1-6 | 63 |
| 4.1.3 | Effects of calyculin A, deltamethrin, and pervanadate | 70 |
| 4.1.4 | Estimating an upper bound on the stoichiometry of phosphorylation | 72 |
| 4.1.5 | Conclusions and discussions..... | 75 |
| 4.2 | Project 2 - Global effects of kinase inhibitors on signaling networks revealed by quantitative phosphoproteomics..... | 78 |
| 4.2.1 | Effects of U0126 and SB202190 on the EGFR signaling pathway | 80 |
| 4.2.2 | Effect of dasatinib on the BCR-ABL signaling pathway | 90 |
| 4.2.3 | Conclusions and discussions..... | 100 |
| 4.3 | Project 3 - ‘Proteomic phenotyping’ to assess differences between transformed and non-transformed mouse liver cells..... | 102 |
| 4.3.1 | Proteomes of primary hepatocytes and Hepa1-6 | 103 |
| 4.3.2 | Phenotyping of the proteomes of primary hepatocytes and Hepa1-6..... | 107 |
| 4.3.3 | Conclusions and discussions..... | 115 |
| 5 | Perspectives | 121 |
| 6 | References..... | 123 |
| | Appendix | 135 |
| | Acknowledgements | 156 |
| | Curriculum Vitae..... | 158 |

Abbreviations

| | |
|-------|--|
| 3D | three dimensional |
| A2M | alpha-2-macroglobulin |
| ABL | Abelson murine leukemia viral oncogene homolog |
| ac | alternating current |
| AGC | automatic gain control |
| AKT | protein kinase B, or Rac (RAC-alpha serine/threonine-protein kinase; the term AKT originates from the transformed AKR mouse strain. |
| ATF2 | activating transcription factor 2 |
| ATP | adenosine triphosphate |
| BCR | breakpoint cluster region protein |
| CAMK | calcium/calmodulin dependent protein kinase; |
| CDKs | cyclin-dependent kinases |
| CID | collision induced fragmentation |
| CK1 | casein protein kinases |
| c-Kit | proto-oncogene tyrosine-protein kinase Kit, or mast/stem cell growth factor receptor |
| CML | chronic myelogenous leukemia |
| dc | direct current |
| DHB | 2,5-dihydroxy benzoic acid |
| DMEs | drug metabolizing enzymes |
| DMSO | dimethyl sulfoxide |
| DNA | deoxyribonucleic acid |
| ECD | electron capture dissociation |
| ECM | extracellular matrix |
| EDTA | ethylenediaminetetraacetic acid |
| EGF | epidermal growth factor |
| EGFR | epidermal growth factor receptor |
| EMA | European Medicines Agency |
| ErbB | epidermal growth factor receptor (EGFR) family |
| ERK | extracellular signal-regulated kinases |
| ESI | electrospray ionization |
| ETD | electro transfer dissociation |

Abbreviations

| | |
|---------------|---|
| FDA | US Food and Drug Administration |
| FDR | false discovery rate |
| Flt-3 | FL cytokine receptor precursor, tyrosine protein kinase receptor |
| FTICR | Fourier transform ion cyclotron resonance |
| GAP | GTP activated protein |
| GIST | gastrointestinal stromal tumors |
| GLEPP1 | glomerular epithelial protein 1 or receptor-type tyrosine-protein phosphatase O |
| GLUT | glucose transporter |
| GO | Gene Ontology |
| GTP | guanosine triphosphate |
| HIFs | hypoxia-induced factors |
| HGF | hepatocyte growth factor |
| HPLC | high performance liquid chromatography |
| HSP | heat shock protein |
| HSP27 | heat shock protein 27 |
| IGF | insulin-like growth factor |
| IMAC | immobilized metal ion affinity chromatography, |
| IP | immunoprecipitation |
| IRMPD | infrared multiphoton dissociation |
| IRS2 | insulin receptor substrate 2 |
| iTRAQ | isobaric tag for relative and absolute quantitation |
| JAKs | Janus kinase, tyrosine protein kinase |
| JNK | C-jun-amino-terminal kinase |
| LDL | low-density lipoprotein |
| LTQ | linear quadrupole ion trap |
| LYN | tyrosine protein kinase LYN |
| m/z | mass to charge ratio |
| MALDI | matrix-assisted laser desorption/ionization |
| MAP3K | mitogen-activated protein kinase kinase kinase |
| MAP2K | mitogen-activated protein kinase kinase |
| MAPK | mitogen-activated protein kinase |
| MAPKAP kinase | MAPK activated protein kinase |
| MBP-C | mannose-binding protein C |

Abbreviations

| | |
|---------------|---|
| MeCN | acetonitrile |
| MEK | dual specificity mitogen-activated protein kinase kinase |
| MS | mass spectrometry |
| MSK2 | nuclear mitogen-and stress-activated protein kinase 2 or ribosomal protein S6 kinase alpha-4 |
| mTOR | mammalian target of rapamycin |
| Myc | Myc proto-oncogene protein, transcription factor p64 |
| nanoESI | nanoelectrospray ionization |
| NF1 | neurofibromin or Neurofibromatosis-related protein NF-1 |
| NF κ B | nuclear factor kappa B |
| p90RSK | ribosomal protein S6 kinase alpha |
| PBS | phosphate buffered saline |
| PCR | polymerase chain reaction |
| PDGFR | platelet-derived growth factor receptor precursor |
| PDK-1 | phosphoinositide-dependant kinase-1 |
| PEP | posterior error probability |
| PI3K | Phosphatidylinositol-4,5-bisphosphate 3-kinase |
| PKA | protein kinase A |
| PLC | phosphoinositide phospholipase C |
| PP1 | protein phosphatase 1, serine/threonine phosphatase |
| PP2A | protein phosphatase 2A, serine/threonine phosphatase |
| PP2B | serine/threonine-protein phosphatase 2B, calmodulin-dependent calcineurin A |
| PTEN | phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase and dual-specificity protein phosphatase |
| PTKs | protein tyrosine kinases |
| PTM | post translational modification |
| PTPs | protein tyrosine phosphatases |
| RasGAP | Ras GTPase-activating protein |
| Rb | retinoblastoma |
| RET | proto-oncogene tyrosine-protein kinase receptor ret precursor |
| rf | radio frequency |
| ROS | reactive oxygen species |
| RP HPLC | reverse phase high performance liquid chromatography |

Abbreviations

| | |
|----------------|--|
| RTKs | receptor tyrosine kinases |
| SCFR | synonym of KIT |
| SCX | strong cation exchange |
| SDS-PAGE | isodium dodecyl sulfate polyacrylamide gel electrophoresis |
| SHIP2 | phosphatidylinositol-3,4,5-trisphosphate 5-phosphatase 2 |
| SHP-1 | tyrosine-protein phosphatase non-receptor type 6 |
| SHP-2 | tyrosine-protein phosphatase non-receptor type 11 |
| SILAC | stable isotope labeling by amino acids in cell culture |
| Smad2/3 | mothers against decapentaplegic homolog 2/3 |
| SOP | standard operating procedures |
| SORI | sustained off resonance irradiation |
| Src | proto-oncogene tyrosine-protein kinase Src |
| STAT | signal transducer and activator of transcription |
| STE | group of homologs of yeast sterile protein kinases |
| TCA | tricarboxylic acid |
| TFA | trifluoroacetic acid |
| TGF- β | transforming growth factor β |
| TGF β R1 | transforming growth factor β type 1 receptor |
| TKL | tyrosine kinase like |
| TNFR | tumor-necrosis factor receptor |
| TSC2 | tuberous sclerosis 2 protein |
| VEGFR | vascular endothelial growth factor receptor |

Summary

This thesis applies quantitative mass spectrometry to research topics in relation to cancer. Proteome-wide quantification at the protein expression level and phosphorylation level were achieved. The technologies developed and used here cover the latest improvements in instrumentation in mass spectrometry, strategies in phosphopeptide enrichment in large scale, algorithms in data analysis and their streamlined implementation, and data mining in downstream bioinformatics. For each of the projects described in this thesis, proteome mapping routinely resulted in identification and quantitation of around 4,000 proteins and phosphoproteome mapping often lead to quantitation of more than 5,000 phosphorylation sites. This ‘systems-wide’ quantitation of the proteome and phosphoproteome is a completely novel development, which has not been used in cancer related topics before.

Three major biology topics are studied in this thesis. In the first project, the phosphoproteome of a mouse liver cancer cell line Hepa1-6 was analyzed in-depth, by using phosphatase inhibitors (calyculin A, deltamethrin, and Na-pervanadate) to boost phosphorylation. The characterization of the phosphoproteome revealed a broad spectrum of cellular compartmentalization and biological functions. Quantitation of phosphatase inhibitor treatment using the Stable Isotope Labeling by Amino Acids in Cell culture (SILAC) method revealed the quantitative effects of these inhibitor compounds on the whole phosphoproteome. To our surprise, these three broadband phosphatase inhibitors displayed very different efficiency, with tyrosine phosphorylation significantly boosted but serine/threonine phosphorylation much less affected. Additionally, a method to estimate an upper bound of the stoichiometry of phosphorylation was introduced by comparing phosphorylation in three SILAC conditions: non-treated cells, stimulated cells (e.g. with insulin), and only phosphatase inhibitor treated cells. The methods developed here can be used directly in development of drugs directed against kinases and phosphatases, key regulators in cancer and other diseases.

The second project continues with the application of phosphoproteomics techniques. Kinase inhibitors influence cellular signal transduction processes and therefore are of great potential in rescuing aberrant cellular signaling in tumors. In fact they constitute a significant portion of drug developing programs in pharmaceutical industry. With the aim of quantifying the effect of kinase inhibitors over the entire signaling network, the second project first set out to study two very commonly used kinase inhibitor compounds for MAPKs: U0126 and SB202190. Their effect on epidermal growth factor (EGF) signal transduction was quantified and

compared using the HeLa cell system. The study confirmed that the MAPK cascades are the predominant signaling branches for propagating the EGF signaling at early time points of stimulation. These large scale examinations also suggest that U0126 and SB202190 are quite specific inhibitors for MAPKs as the majority of regulated phosphopeptides appears to belong to the MAPK pathways. In the second part of the project, the effect on phosphoproteome changes of the chemical compound dasatinib, which was demonstrated to effectively inhibit the constitutively activated fusion protein BCR-ABL and was recently approved for chronic myelogenous leukemia (CML) therapy, was quantified in the human CML cell line K562. Bioinformatic analysis revealed that the most influenced signal transduction branch was the Erk1/2 cascade. Overall more than 500 phosphorylation sites were found to be regulated by dasatinib, the vast majority not described in the literature yet.

The third project compared the proteomes of mouse hepatoma cell line Hepa1-6 with the non-transformed mouse primary hepatocytes. This was performed by combining the SILAC heavy labeled form of Hepa1-6 with the primary hepatocytes. To characterize the features of these two proteomes, quantitation information (i.e. protein ratios between the two cell types) was used to divide all proteins into five quantiles. Each quantile was clustered according to the Gene Ontology and KEGG pathway databases to assess their enriched functional groups and signaling pathways. To integrate this information at a higher level, hierarchical clustering based on the p-value from the first Gene Ontology and KEGG clustering was performed. Using this improved bioinformatic algorithm for data mining, the proteomic phenotypes of the primary cells and transformed cells are immediately apparent. Primary hepatocytes are enriched in mitochondrial functions such as metabolic regulation and detoxification, as well as liver functions with tissue context such as secretion of plasma and low-density lipoprotein (LDL). In contrast, the transformed cancer cell line Hepa1-6 is enriched in cell cycle and growth functions. Interestingly, several aspects of the molecular basis of the “Warburg effect” described in many cancer cells became apparent in Hepa1-6, such as increased expression of glycolysis markers and decreased expression of markers for tricarboxylic acid (TCA) cycle.

Studies in this thesis only provide examples of the application of mass spectrometry-based quantitative proteomics and phosphoproteomics in cancer research. The connection to clinical research, especially the assessment of drug effects on a proteome wide scale, is a specific feature of this thesis. Although this development is only in its infancy, it reflects a trend in the quantitative mass spectrometry field. We believe that more and more clinical related topics can and will be studied by these powerful methods.

1 Introduction

In this proteomics era, mass spectrometry (MS) is a powerful technology to explore important questions in biology. Recent progress in MS has enabled close-to-complete proteome measurements¹. Quantity comparison has become very precise even without the aid of chemical or metabolic labeling. These achievements have encouraged researchers to apply MS to various topics in biology and medicine, including cancer research.

1.1 Signal transduction in cancer

Signal transduction is an inherent molecular mechanism to respond to extracellular environmental changes and to maintain cellular homeostasis. Growth factor signaling is a canonical paradigm for response to extracellular signals and such processes are involved in growth², differentiation^{3,4}, development⁵⁻⁷, immune response⁸, and cell movement⁹, to name a few examples. Signaling involved in cellular homeostasis coordinates intracellular organization, ensures survival of normal somatic cells^{10,11} and removes senescent or damaged cells¹², among many other functions.

1.1.1 Signal transduction paradigm

Signal transduction is a triggered event. It typically consists of three elements: the activator which initiates the signal transduction, the effectors which propagate the signal, and the attenuators which provide negative feedback and eliminate the signal.

Epidermal growth factor receptor (EGFR) signaling is a canonical signal transduction pathway (Figure 1.1). Ligand binding to the extracellular domain induces EGFR dimerization and conformational change. The exposed tyrosine residues of the intracellular kinase domain in one monomer are cross-phosphorylated by the other monomer. Phosphotyrosines create binding sites and thus bring to the intracellular domain of EGFR numerous SH2 domain containing proteins, such as RasGAP, SHP-1, and PI3K. In turn, these proteins recruit more signaling molecules to propagate the signal. As a balance, negative feedback loops are initiated to attenuate and terminate the signal^{13,14}.

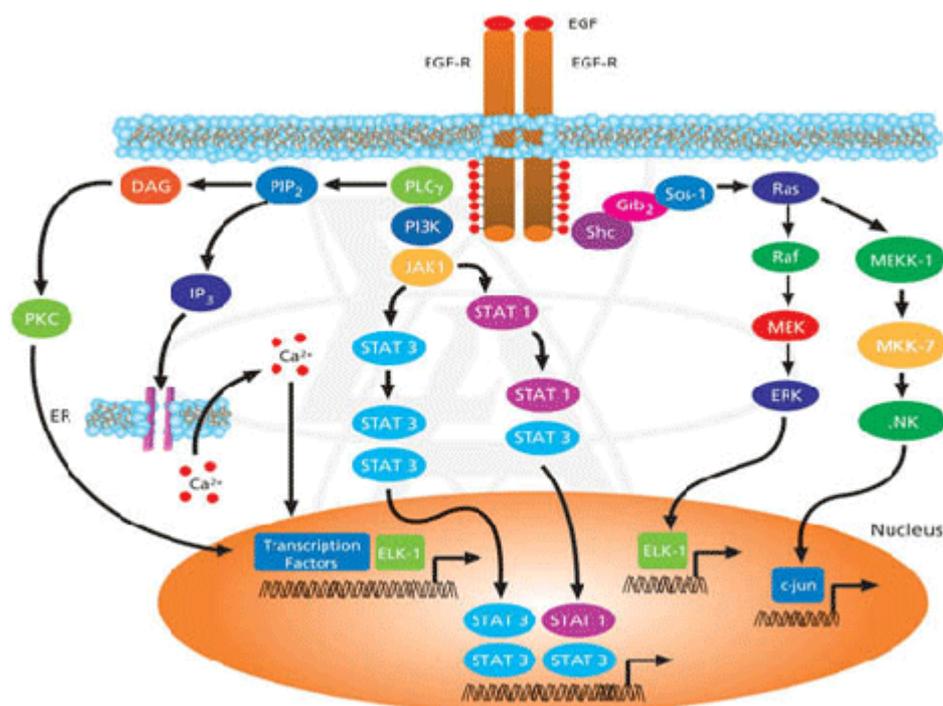


Figure 1.1 A canonical growth factor signal transduction network – the EGFR signaling pathway. From Sigma Aldrich Key Resources.

It has been estimated that up to 20% of the human genome encodes signaling proteins^{15, 16}. These include 1,543 signaling receptors¹⁷, 518 protein kinases¹⁸, around 150 protein phosphatases¹⁹⁻²¹, and more than 1,850 transcription factors^{17, 22}.

1.1.2 Mechanism of signal transduction

Actions such as conformational change, post-translational modifications, molecular interactions, regulated localization, and trafficking are essential to signal transduction. It typically is an amplifying process where one upstream effector reacts with multiple downstream effectors.

Phosphorylation is by far the most well studied mechanism of regulation in signal transduction. It is also the major subject of this thesis. However, it is noteworthy that research in recent years has revealed more and more post-translational modifications as important regulatory mechanisms. For example, ubiquitination – the addition of ubiquitin-like molecules to lysine residues – is widely recognized as governing protein degradation via the “ubiquitin-

proteasome pathway". It also plays a role in protein trafficking, DNA repair and replication, and other cellular functions. A related small protein modification – SUMOylation – has been demonstrated to be involved in processes like cell cycle progression and apoptosis. As another example of an important PTM, acetylation of lysine residue in histone tails generally leads to active gene transcription.

To ensure speed, specificity and efficiency in signal transduction, several mechanisms are employed.

Domain-domain interactions. Structurally conserved protein domains provide recognition patterns for biomolecular interactions. Currently the PFAM database documents more than 3500 domains for *Homo sapiens*²³ and a subset of these are commonly employed in signal transduction²⁴. Signaling proteins may contain tandem domains or multiple types of domains to facilitate their functions.

Different dynamics. Reaction dynamics vary for different proteins^{25, 26} and protein groups^{27, 28}. Compared to long-term stimulation, signals lasting for only a few seconds can trigger different reaction modes^{27, 29}. Often, similar dynamics are observed among correlating proteins, such as components of complexes^{30, 31}. Cells use different signal attenuating strategies. While Ser/Thr phosphorylation and tyrosine dephosphorylation are often used to quickly attenuate the signal³²⁻³⁴, other long term negative feedbacks are also employed, such as transcriptional and translational control, protein turnover, and exocytosis^{35, 36}.

Compartmentalization and scaffolding. Co-localization concentrates signaling proteins and therefore enhances interaction efficiency³⁷. For example, in response to insulin, most proteins involved in GLUT4-mediated glucose transport gather in caveolar lipid rafts³⁸. Multiple localizations may indicate alternative functions of the same protein. For example, the receptor tyrosine kinase Flt3 localized on organelle membranes has different activity from that on the plasma membrane³⁰.

Tissue specific expressions and isoforms. Proteins can be selectively expressed according to their function in different tissues and cell types. For example, GLUT4 is expressed predominantly in insulin responsive tissues^{25, 39}. It has been estimated that around 40-60% of human genes are subjected to alternative-splicing events^{40, 41}. Isoforms may perform different functions. For example, PLC γ is the only PLC isoform that contains signal transduction domains, indicating its active role in signal transduction⁴².

1.1.3 Altered signal transduction in cancer

In the homeostatic state, signal transduction pathways form a well coordinated and subtly balanced network. However, this balance is distorted in tumors, where cell proliferation and survival are strengthened and apoptosis is impaired. Cancer, a malignant form of neoplasm, is able to invade normal tissues and metastasize to and grow at distant body sites⁴³.

Cancer is characterized by self-sufficiency in growth signals, insensitivity to growth-inhibitory signals, evasion of apoptosis, limitless replicative potential, tissue invasion and metastasis, and sustained angiogenesis^{44, 45}. It is an intractable disease with multiple origins and complicated etiology.

1.3.3.1 Relentless cell proliferation in cancer

Relentless cell proliferation in cancer lesions is a prerequisite for cancer transformation⁴⁶. There are several check points that guard cell cycle progression. The final commitment to cell duplication is the transition from G1 phase to S phase. In this process, the retinoblastoma protein (pRb) plays a critical role. Inactivation or deletion of pRb can be found in most human cancers^{47, 48}. Likewise, p53, the ‘guardian of the genome’, is mutated in most tumors as well as Myc, a transcription factor and another critical player in cell cycle^{47, 48}.

CDKs, which are Ser/Thr protein kinases, are the central engines that propel the cell cycle. In contrast, inhibitory proteins such as p21^{Waf1/Cip1} and p27^{Kip1} negatively regulate this process. In cancer, enhanced activities of CDKs and loss-of-function of p21^{Waf1/Cip1} and p27^{Kip1} are often reported⁴⁹.

1.1.3.2 Enhanced cell growth in cancer

Aberrations in three signaling cascades, Ras, PI3K and mTOR, are considered critical in promoting tumor cell growth⁵⁰. Mutations in the master regulators, such as K-Ras, H-Ras, N-Ras, and the p110 α catalytic subunit of PI3K, are found in most human tumors⁵⁰. B-Raf activity is enhanced in more than 60% of human malignant melanomas, as well as in some colon, thyroid and lung tumors⁵¹. Furthermore, mutation in one or another PI3K pathway component accounts for up to 30% of all human cancers⁵². Mutations in the negative feedback loops also contribute to cancer development. For example, inactivation of the GTPase activating protein NF1 leads to the accumulation of GTP-bound Ras, and therefore promotes

Ras signaling. Sporadic mutations of PTEN, a lipid phosphatase that balances PI3K signaling, are so frequently found in tumors that they are now regarded as the second commonly mutated tumor suppressor in humans, right after p53⁵⁰.

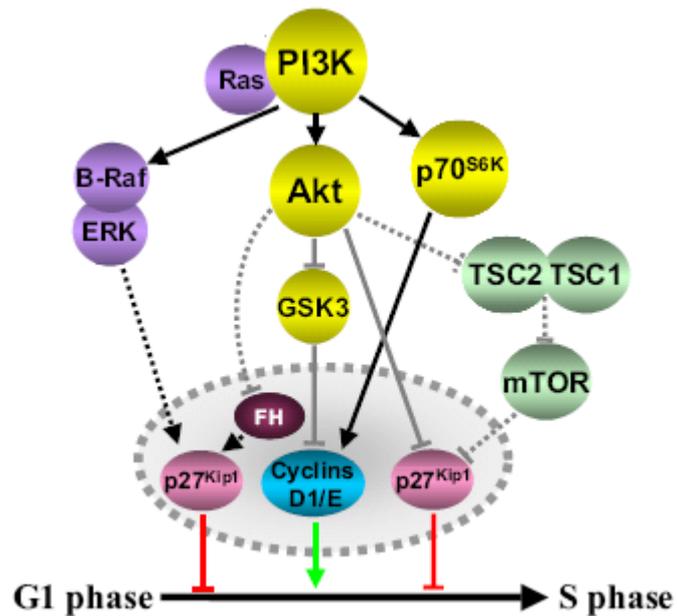


Figure 1.2 The Ras, PI3K, and mTOR activate cell cycle by inhibiting CDK inhibitors (e.g. p27^{Kip1} and p21^{Waf1/Cip1}) and activating cyclins. Modified from New *et al.*⁵³.

The mechanisms of the Ras, PI3K and mTOR pathways to enhance tumor growth are manifold. First, they activate CDKs and inhibit cell cycle suppressors, thereby drastically promoting cell proliferation^{53, 54} (Figure 1.2).

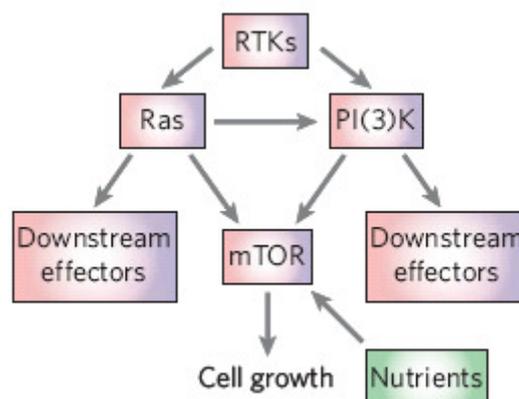


Figure 1.3 Connection of the three major signaling cascades in promoting neoplasm: Ras, PI3K, and mTOR signaling cascades. From Shaw *et al.*⁵⁰.

Second, constitutively activated signaling can bypass survival factors, which are required for the survival of normal cells. Third, hyper-activated mTOR signaling promotes protein synthesis and angiogenesis even under nutrient-deficient conditions⁵⁵. This ensures synthesis of the protein material for tumor cell expansion. Strikingly, these three pathways are tightly interconnected (Figure 1.3). They work in concert to promote tumor growth.

1.1.3.3 Suppressed apoptosis in cancer

Apoptosis or programmed cell death is mainly implemented via caspases⁴⁶. They can be activated by the ligands of the death receptors or by cytochrome c, which is released from mitochondria. In tumors, mutations are reported to suppress apoptosis, for example by impairing the release of cytochrome c and therefore violate the major initiating mechanism of apoptosis⁵⁶⁻⁵⁹. The transcription factor p53 plays a critical role in apoptosis. It induces expression of the cell cycle inhibitor p21^{waf/cip} and the pro-apoptotic protein Bax⁴² and also functions as an exonuclease in DNA damage repair responses⁶⁰.

1.1.3.4 Other critical signaling pathways in cancer development

Besides the aforementioned signaling pathways, other critical events in cancer development have become increasingly clear. The NF κ B signaling pathway bridges cancer with infection and chronic inflammation⁴⁵. Hypoxia induced factors (HIFs) trigger angiogenesis under the hypoxia conditions typical of the cancerous microenvironment⁵⁵. A number of signaling factors, including TGF- β , HGF and IGF, can facilitate metastasis⁶¹.

1.1.4 Kinases, phosphatases and cancer

1.1.4.1 Families of protein kinases and protein phosphatases

Phosphorylation and dephosphorylation are essential for most signal transduction events. Sequencing of the human genome and sequence comparison revealed 518 genes encoding protein kinases¹⁸ and 122 genes encoding protein phosphatases^{21, 62}.

Protein kinases are among the largest protein families in eukaryotes. By sequence homology of their catalytic domains, they can be classified into 9 broad groups, 90 families and 145 subfamilies¹⁸ (Figure 1.4 and Appendix 1). Based on substrate specificity, protein kinases can

Introduction

be divided into Ser/Thr kinases, tyrosine kinases (PTKs), and dual-specificity kinases. While the former two types of kinases selectively phosphorylate serine/threonine or tyrosine residues, respectively, the dual-specificity kinases can phosphorylate all of them. Large scale phosphoproteomics studies reestablished that tyrosine phosphorylation only accounts for a small percentage of overall phosphorylation, with a ratio of pSer:pThr:pTyr of 48:7:1³¹. Out of the 518 genes, only 90 encode tyrosine kinases.

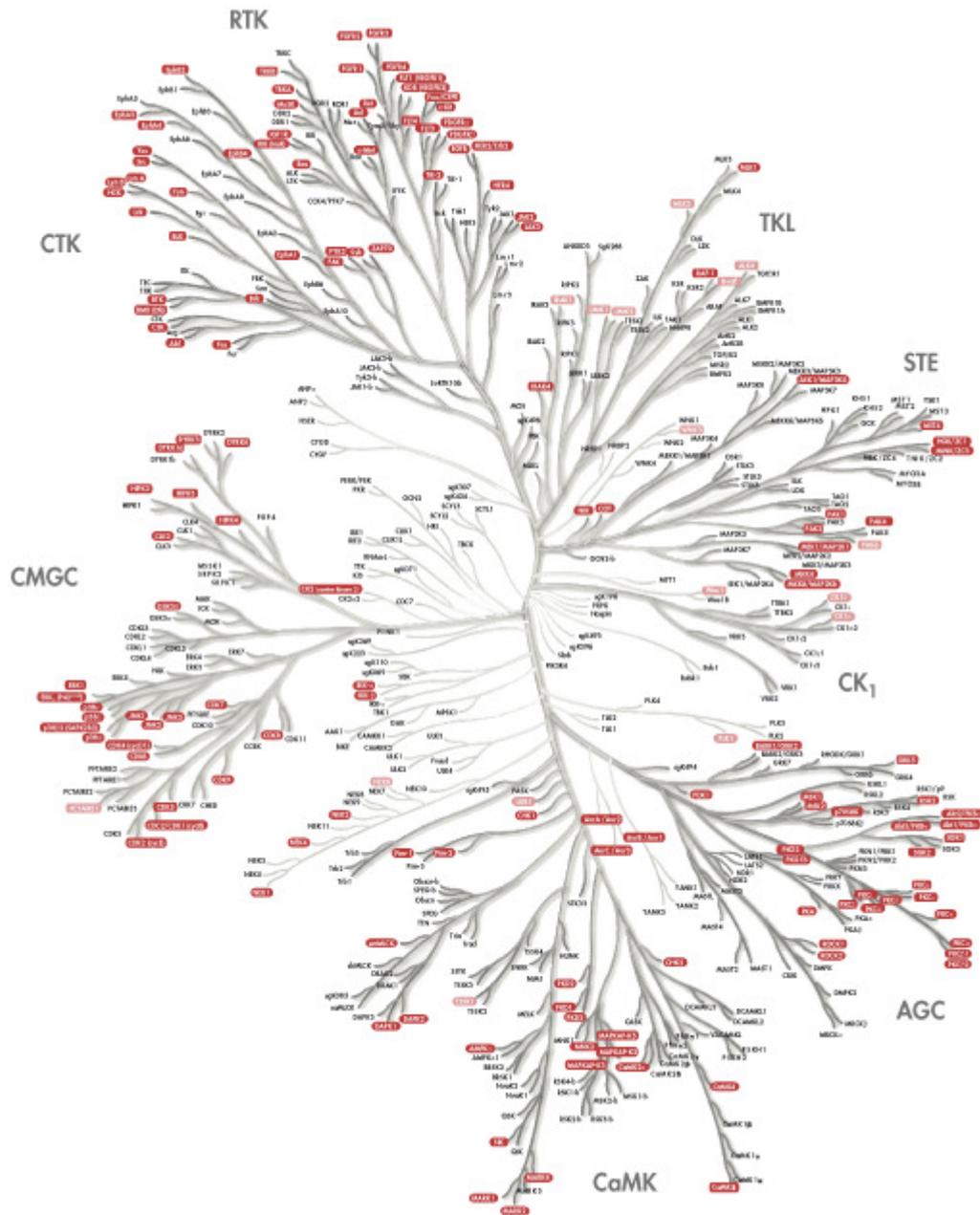


Figure 1.4 Dendrogram of the human kinome. **TK**: tyrosine kinase group; **TKL**: tyrosine kinase like group; **STE**: group of homologs of yeast sterile protein kinases; **CK1**: casein

protein kinases; **AGC**: PKA, PKG & PKC containing group; **CAMK**: calcium/calmodulin dependent protein kinase group; **CMGC**: CDK, MAP, GSK3 & CLK containing group. Adapted from: Cell Signaling Technology.

Protein phosphatases counteract and balance the action of protein kinases. They are generally divided into Ser/Thr phosphatases and tyrosine phosphatases (PTPs) (Figure 1.5). Interestingly, there are 107 genes encoding tyrosine phosphatases²¹ while only 15 genes encode Ser/Thr phosphatases⁶². This suggests that dephosphorylation is regulated differently for tyrosine and serine/threonine. Indeed, Ser/Thr phosphatases often recruit distinct regulatory proteins to achieve substrate specificity. In contrast, most PTPs possess multiple domains which mediate their specificity towards substrates.

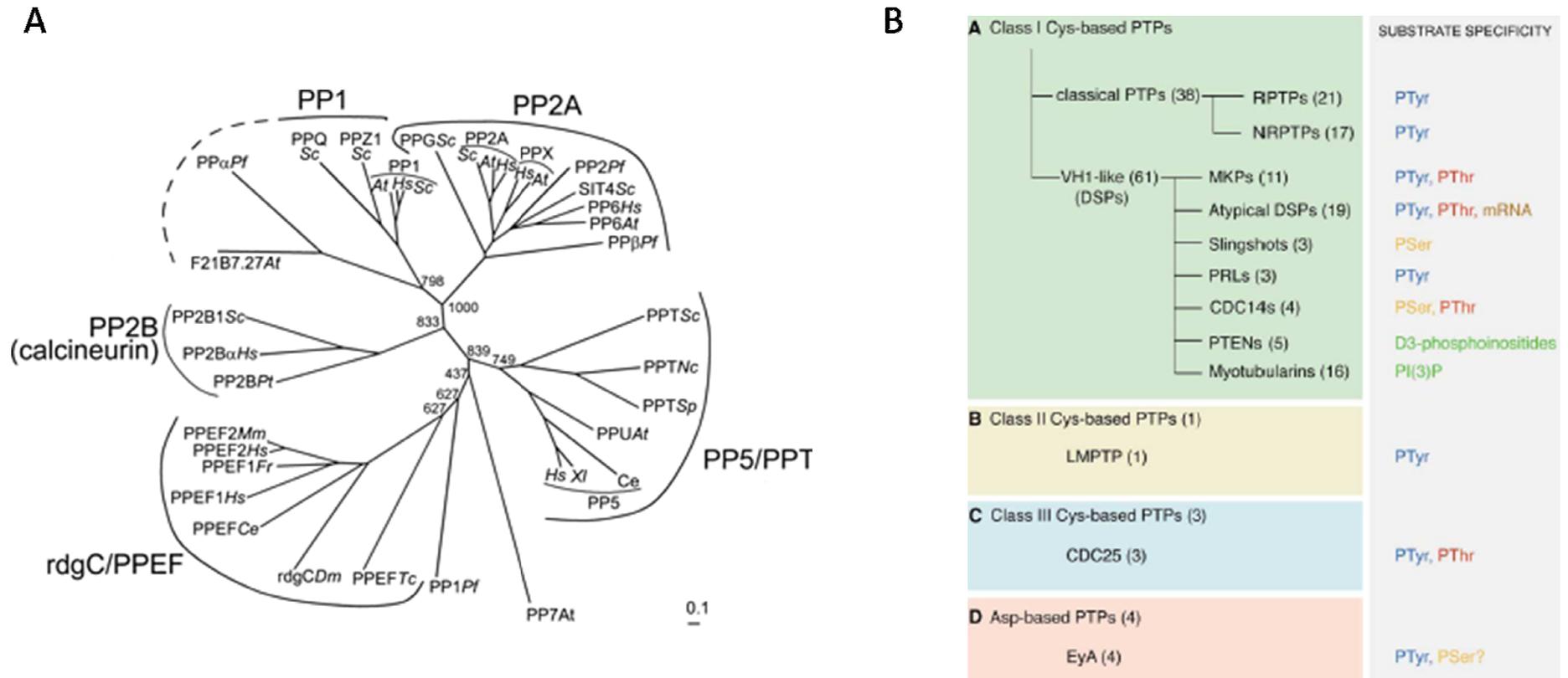


Figure 1.5 Phosphatase superfamily. (A) Dendrogram of the serine/threonine protein phosphatases, from Andreeva *et al.*⁶³. (B) Classification of the tyrosine protein phosphatases, from Alonso *et al.*²¹.

1.1.4.2 Protein kinases and protein phosphatases in cancer

Over 400 human diseases have been connected to protein kinases. In fact, protein kinases especially PTKs, comprise a large fraction of the more than 100 dominant oncogenes known to date. Aberrations of the 31 receptor tyrosine kinases (RTKs) are frequently found in human cancers¹⁶. These include Neu/ErbB2 and EGFR in breast and lung carcinomas and Kit/SCFR in several highly malignant tumors such as gastrointestinal stromal tumor, acute myeloid leukemia, and lung cancer. Aberrations of cytoplasmic PTKs, such as c-Src, c-Abl, JAKs and STATs, are also associated with human malignancy.

Protein phosphatases generally function as negative regulators of signal transduction. Therefore, abnormalities of PTPs often promote cancer development. For example, point mutations are commonly observed. In one report, six mutated PTPs were found in colon cancer, with equal distribution in receptor and non-receptor PTP families³⁴. Strikingly, these mutations were mostly located outside the catalytic domains. Another format of abnormality is hyper-methylation of the genes for PTPs. Well documented examples include hyper-methylation-mediated silencing of *shp1* and *glepp1*⁶⁴⁻⁶⁷. Furthermore, those PTPs which can positively regulate signal transduction are also reported to malfunction in malignancy. For example, SHP2 is reported to be hyper-active in Noonan syndrome and leukemia^{68, 69}.

1.1.5 Molecular targeted cancer therapies

The heavy reliance of many cancers on critical proteins, especially kinases and phosphatases, has stimulated interest in developing molecular targeted therapies. Clinical demonstration of this concept was only demonstrated at the turn of the 21st century, with the introduction of Gleevec (see below). To date, a decade's development has resulted in the successful introduction of several inhibitor compounds and antibodies into clinical use. In fact, this exciting process is full of reversals, re-discoveries and innovations⁵⁴.

1.1.5.1 Small molecule inhibitors for kinases and phosphatases

In the clinic, cyclosporine A has been used as an immunosuppressant for organ transplantation surgeries since 1983. It was not until 1991 that its molecular mechanism was revealed – it is an inhibitor for phosphatase calcineurin (PP2B)⁷⁰. Rapamycin, another immunosuppressant approved for clinical use in 1999, targeted mTOR with the highest specificity among all

inhibitors to date⁷¹. Fasudil, which mainly inhibits the Rho-dependent protein kinase Rock, was approved in Japan in 1995 for the treatment of cerebral vasospasm⁷¹. Certain indirubin dyes, included in an ancient Chinese herbal remedy for treating leukemia for hundreds of years, were discovered to be potent CDK inhibitors⁷².

The examples mentioned above indicate that some kinase or phosphatase inhibitor drugs were in use long before the concept of molecular targeted inhibition was articulated. Their success alleviated two initial concerns, that inhibitors may not achieve sufficient specificity due to the conservation of the catalytic domain in protein kinases, and that they may not be able to compete with the highly concentrated cellular ATP (2-10 mM).

The first inhibitor drug developed with the clear aim of molecular targeting to a specific cancer protein is imatinib. It entered human clinical trial in 1996 and was approved for clinical use in 2001⁵⁴. The molecular target of imatinib (Gleevec) is BCR-ABL and the drug achieved very high remission rates. It also inhibits the PDGF receptor and cKIT and achieved 80% response in gastrointestinal stromal tumors patients with activating KIT mutations⁷³. With its high efficacy, minimum side effects, and short time to approval, imatinib has propelled rapid development of other inhibitor compounds. After G-protein-coupled receptors, protein kinases are now the second largest group of drug targets in the pharmaceutical industry, comprising 20-30% of drug development programs. An up-to-date list of inhibitor compounds in clinical use is displayed in Table 1.1. Over 75 protein kinase inhibitors have entered advanced human clinical trials, and it is estimated that over 500 other inhibitor compounds are in preclinical studies (http://www.kinexus.ca/science/protein_kinases/).

1.1.5.2 Specificity of the small molecule inhibitors

Most kinase inhibitor compounds are ATP competitive. Many of them are known to target more than one kinase⁷¹. However, those that achieve high specificity often interact with neighboring regions of the ATP binding pocket. This is well demonstrated by the crystal structures of SB203580 binding to p38 MAPK and purvalanol binding to CDK2⁵⁴. Crystal structures also revealed that imatinib straddles the highly conserved amino-terminal region of the activation loop in BCR-ABL and locks the kinase in its inactive conformation⁷⁴. The compound BIRB0796 binds to p38 MAPK in a similar way⁵⁴.

Table 1.1 Kinase inhibitor compounds in clinical use or clinical trials. From Petrelli *et al.*⁷⁵.

| Single target drugs | Targets | Clinical use |
|---------------------------|---|---|
| Trastuzumab | HER2 | HER-2 over-expressing breast cancer |
| Cetuximab | HER1 | CRC, HNSCC |
| Bevacizumab | VEGF | CRC, NSCLC |
| Pertuzumab | HER2 | Phase II trials: breast, ovarian cancers, NSCLC |
| Gefitinib | HER1 | NSCLC |
| Erlotinib | HER1 | NSCLC |
| Multi target drugs | | |
| Imatinib | BCR-Abl, c-Kit, PDGFR | CML, GIST |
| Sunitinib | VEGFR, PDFGR*, c-Kit*, Flt-3, RET, CSF-1R | RCC, imatinib-resistant GIST |
| Sorafenib | b-RAF, VEGFR, PDFGR, c-Kit, Flt-3 | RCC |
| Lapatinib | HER1, HER2 | Phase II-III: breast cancer, NSCLC |
| Vandetanib | HER1, VEGFR, RET | Phase II-III: NSCLC, thyroid cancer |

*also imatinib-resistant mutated forms.

CRC, colorectal cancer; HNSCC, head and neck squamous cell carcinoma; NSCLC, non-small-cell lung cancer; CML, chronic myelogenous leukemia; GIST, gastrointestinal stromal tumor; RCC, renal cell cancer.

1.1.5.3 Antibody-directed therapy

Molecular targeted therapies also include other strategies beyond kinase and phosphatase inhibitors. Antibody therapy is one of these and increasing in popularity. Antibodies that bind to the extracellular domain of RTKs can correct their hyper-active signaling, which is due to over-expression or mutation. For example, cetuximab is a monoclonal antibody for EGFR which recently received both the FDA (the US Food and Drug Administration) and EMEA (the European Medicines Agency) approval to treat head and neck cancers⁷⁶. Bevacizumab, a monoclonal antibody against VEGF, has been approved by the FDA as first line cancer therapy that acts against angiogenesis⁷⁷.

1.1.5.4 Combined and multi-target therapies

Relapse after kinase inhibitor treatment is frequently seen in the clinic and this has encouraged the development of second generation inhibitors⁷⁸. So far, cancer cells are known to engage in at least two strategies to escape the drug's effects: to generate second mutations in the targeted kinases^{79, 80} or to use other signaling molecules to compensate for the suppressed kinases^{81, 82}. As a solution, it is suggested to rationally combine inhibitor drugs that can target multiple oncogenes in parallel. For example, combination of imatinib and dasatinib for treating chronic myelogenous leukemia (CML) has been proposed⁸³⁻⁸⁵.

1.2 Mass spectrometry in biology

Nearly 20 years after its introduction to biology research, MS has become a tremendous success⁸⁶. It has revolutionized the way in which biological information, especially related to proteins, can be obtained. With time, MS will become a routine technique to tackle a wide variety of biological questions.

1.2.1 General workflow of proteomics

Mass spectrometry measures the mass to charge ratio (m/z) of molecules. In MS-based proteomics, the m/z values of peptides or small proteins are measured, which reflects their amino acid composition and possible post-translational modifications. Generally there are two measuring approaches, top-down and bottom-up.

Bottom-up is the most widespread and by far the most successful method in proteomics (Figure 1.6). Proteins are enzymatically cleaved at specific sites to yield short peptides which typically consist of 6-20 amino acid residues. Mass spectrometers firstly examine the m/z value of these peptides (survey scan or MS¹ scan) and secondly measure their fragments generated inside the mass spectrometer by low energy excitation (MS/MS fragmentation). Taking a protein database as a reference, mass spectra can be correlated to amino acid sequences with the aid of computer algorithms. The found peptide sequences are then assigned to proteins, which ultimate leads to protein identification⁸⁷.

In the top-down approach, small intact proteins are ionized and sprayed into mass spectrometers where peptide fragments are subsequently generated using one of a variety of activation methods such as CID, ECD, SORI, and IRMPD methods (see abbreviation page). Top-down proteomics provides an alternative approach in proteomics to peptide based approaches. It obtains better sequence coverage of protein identifications and benefits post translational modification (PTM) research. However, it is usually limited to small proteins with molecular weight up to 30 kDa⁸⁸ and suffers from the difficulty to solubilize and separate proteins in the same way as can be done for peptides.

Proteomics often deals with complex protein or peptide mixtures. As dynamic range and sequencing speed are the limiting factors in the current MS technology²⁸, sample complexity has to be reduced. On-line separation with reverse phase chromatography (HPLC) connecting

on-line to the mass spectrometer has proven to be a useful separation method (Figure 1.6). The C₁₈ reverse phase HPLC column elutes peptide mixtures with linearly increasing organic solvent, e.g. acetonitril (MeCN). The gradual elution (typically at a few hundred nanoliter per minute) with shallow gradients increases available sequencing time in the MS. Prior to this hydrophobicity-based peptide separation, complex samples can be first be separated by one-dimensional gel⁸⁷, isoelectric focusing⁸⁹, ion-exchange^{90, 91}, molecular size⁹², and affinity binding such as immunoprecipitation^{27, 93}, IMAC^{94, 95} and TiO₂ enrichment^{28, 96, 97}.

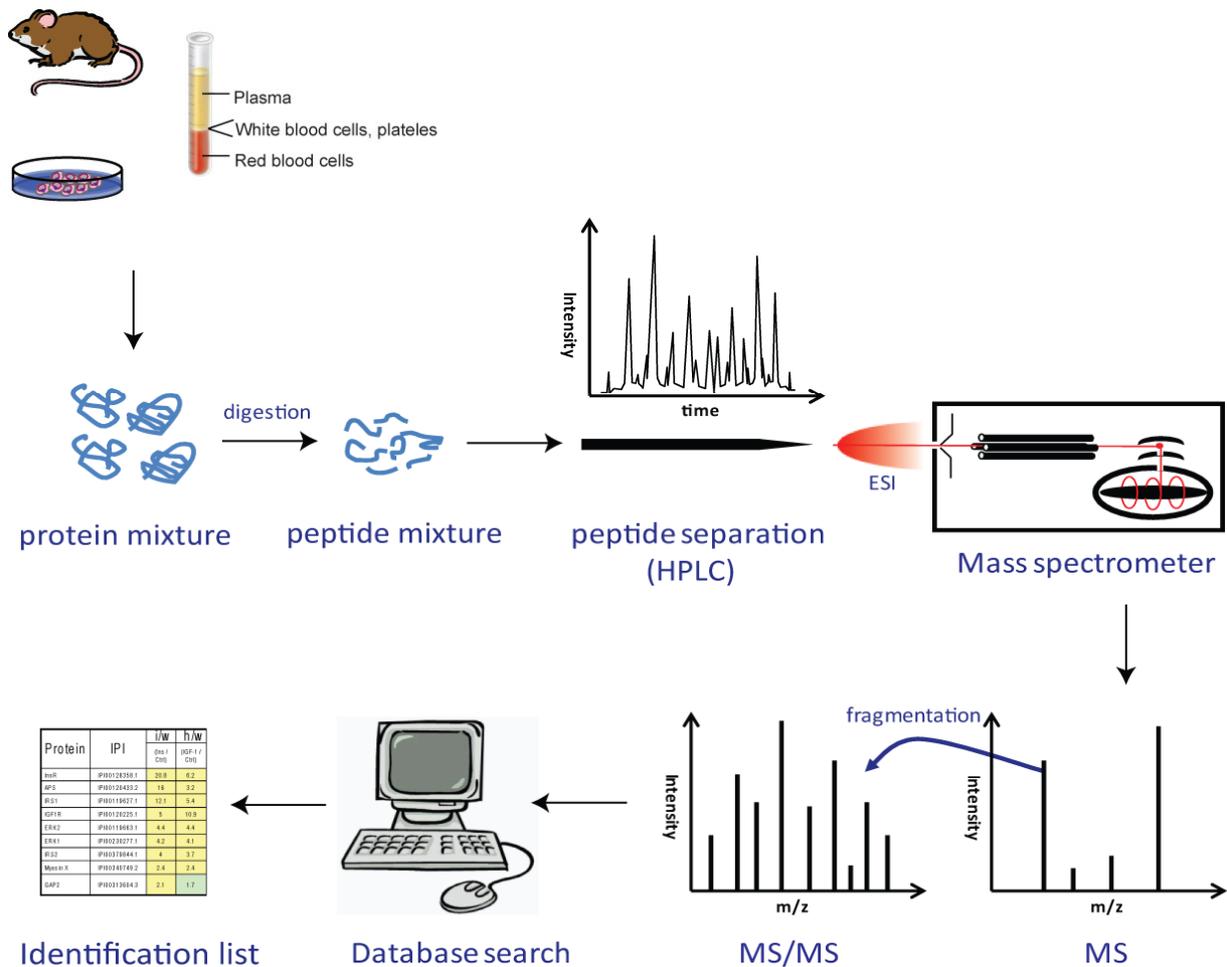


Figure 1.6 General workflow for bottom-up MS-based proteomics. Proteomics samples come from tissues, cell lines, body fluids, etc. Protein or peptide samples can be fractionated by different means to reduce complexity. Depicted is the nanoLC-MS/MS mode, where HPLC is coupled to MS via an electrospray ion source and MS/MS fragmentation is performed to generate information of peptide primary structure. Computer algorithms match mass spectra to amino acid sequences. The outcome of the experiment is a list of identified proteins.

For less complex samples, static electrospray – called nanoelectrospray – can be used which often yields better identification results. In particular static spray is beneficial for some modification studies⁹⁸.

Data analysis and information mining takes place at the end of the work flow. For example, much biological information can be discovered by using different clustering algorithms. Frequently used public resources are the Gene Ontology (GO; <http://www.geneontology.org/>) and KEGG pathway databases (<http://www.genome.jp/kegg/>). Integration of MS-based proteomics with other “omics” datasets can provide deeper insights. Examples of these valuable datasets are microarray based transcriptome studies⁹⁹ and protein-protein interactome and protein complex studies^{100, 101}. Modeling of molecular networking is emerging. It is expected that integration of these large-scale studies will deliver a new kind of biology knowledge that cannot be obtained by each of the separate approaches.

1.2.2 Mass spectrometric instrumentation

Every mass spectrometer consists of an ion source, a mass analyzer, and a detector. Peptide/protein ions are generated in ion source typically by one of two soft ionization methods: Matrix-assisted laser desorption/ionization (MALDI)¹⁰², and electrospray ionization (ESI)¹⁰³. Inside mass spectrometers, peptides ions can be separated according to their momentum in magnetic sector^{104, 105}, kinetic energy in electrostatic sector¹⁰⁶, velocity in time-of-flight instruments¹⁰⁷, path stability in linear quadrupoles¹⁰⁸, as well as orbital frequency in quadrupole ion traps¹⁰⁹, ion cyclotron resonance mass spectrometer^{110, 111}, and Orbitrap mass spectrometer¹¹². Finally, ions are detected under high vacuum conditions in detectors. Often the electron signals are multiplied via secondary electron multiplier.

Standard parameters to evaluate a mass spectrometer include resolution, mass accuracy, mass range or upper mass limit, and ion dynamic range¹¹³.

1.2.2.1 Electrospray ionization

The idea of using electrospray dispersion to produce gas phase ions from solution was first introduced by Dole and colleagues in 1968¹¹⁴. Fenn and coworkers realized this idea and developed the modern day technique of electrospray ionization mass spectrometry (ESI MS)¹¹⁵. Further improvement by Mann and coworkers introduced nanoelectrospray ionization

(nanoESI) which is compatible with nano-flow rates (i.e. at the scale of nanoliter-per-minute) for minute amount of samples^{116, 117}.

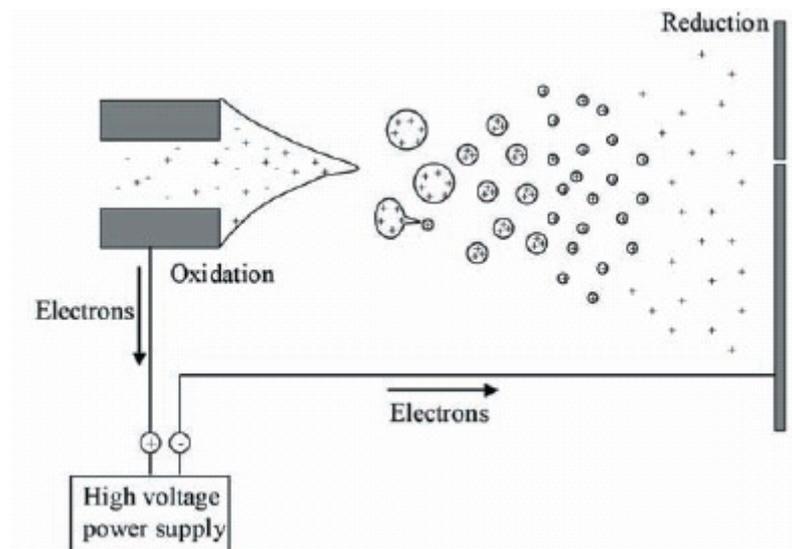


Figure 1.7 Electro spray ionization (ESI) process. From Nielsen¹¹⁸.

In nanoLC ESI (Figure 1.7), a high voltage of several kilovolts is applied to the thin, sharp tip of a metal capillary towards the ion source of the mass spectrometer. The orifice of the ion source is warmed up to around 150-200 degrees. Liquid flows through the capillary to the tip at a slow flow rate, submicroliter-per-minute in typical set ups. The emerging liquid surface disintegrates at such high voltages and temperature condition and liquid droplets are formed. Because they have very high surface to volume ratios, droplets quickly shrink by evaporation. As solutes are of the same charge, electrostatic repulsion leads to explosion of the liquid droplets, a phenomenon termed Coulombic fission. Eventually the strong Coulombic repulsion overcomes the surface tension of the droplet, and ions are liberated into the gas phase and enter the mass spectrometer as single ions¹¹⁹.

1.2.2.2 Linear quadrupole ion trap

Principles of the LTQ

The linear quadrupole ion trap (here the LTQ from Thermo Fisher Scientific) contains two pairs of orthogonally positioned hyperbolic rods, each segmented into three sections (Figure

1.8). The central section is about three times larger than the two end sections. Small slits which cut along the rods at the Z axis provide the ion exits. Therefore two detectors are placed symmetrically next to the slits to record the radially ejected ions¹²⁰. Direct current (dc) voltages are applied to the opposing rods. These opposing rod pairs receive the same voltage, while voltages of the neighboring rods are opposite but of the same amplitude. To trap ions in the axial direction (Z axis), different dc voltages are applied to the three sections to create a deep electric potential well in the center section. To trap ions in the radial direction (XY-plane), the major radio frequency (rf) is applied to the rod pairs at both x and y axis, with the same amplitude but opposite phases, e.g. $V_{rf} \cos\omega t$ and $-V_{rf} \cos\omega t$. To assist ion activation, isolation, and ejection, two phases of supplemental alternating current (ac) voltage are imposed on the X-electrode.

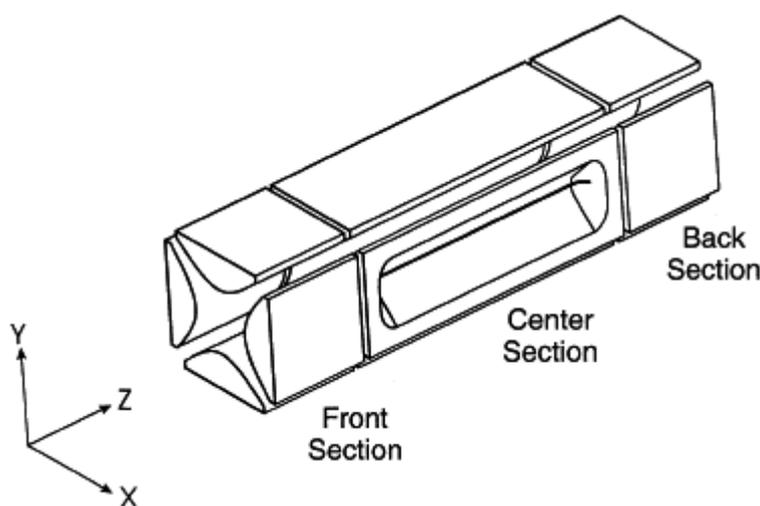


Figure 1.8 Structure of the two-dimensional linear ion trap. From Schwartz *et al.*¹²⁰.

Independent of kinetic energy and initial position, ions obtain stable trajectories within the LTQ only if their m/z values locate in certain stable regions in the ion stability diagram¹²¹ (Figure 1.9). For practical reasons, the region A is most interesting, where q_{\max} equals 0.908. As seen in equation (1), the m/z value of an ion is inversely proportional to the Mathieu parameter q_z . From this the minimum m/z values that can be detected in the LTQ is determined. The same principle also applies to the three dimensional (3D) ion trap, common before the introduction of this 2D ion trap LTQ. This low mass cut-off phenomenon is termed “1/3 cut-off” rule for ion traps¹²².

$$(1) \quad \frac{m}{z} = \frac{4eV}{q_z \omega^2 r_0^2}$$

$$(2) \quad a_z = \frac{-8zeU}{m\omega^2 r_0^2}$$

$$(3) \quad q_z = \frac{4zeV}{m\omega^2 r_0^2}$$

Where m is the ion mass; z is the ion charge; U and V are the potentials of dc and rf, respectively; a_z and q_z are Mathieu parameters; ω is the angular frequency of the rf; r_0 is the radius inscribed, i.e. one-half distance between the opposite rods.

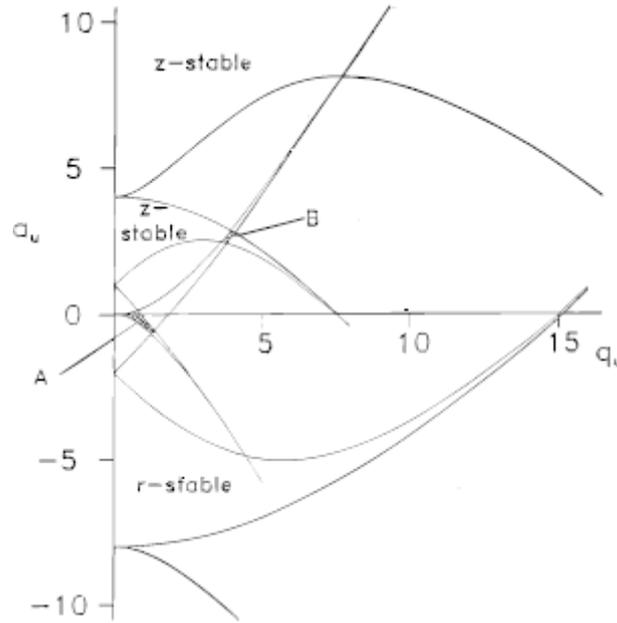


Figure 1.9 The ion stability diagram of quadrupole ion trap. Mathieu parameters (a_u , q_u) are plotted to identify the stability regions in both the XY- (radial) and Z- (axial) directions. Regions of simultaneous overlap are marked as A and B. From March *et al.*¹⁰⁹. The Mathieu parameters can be calculated by equation (2) and (3).

To analyze trapped ions, the principle of mass-selective instability is employed where the main rf voltage is increased at a constant rate. The ramping of the rf potential amplitude causes the q_z values of all ion species to increase and eventually to exceed the resonance

ejection limit (set as 0.88 in some cases¹²⁰). Ions are radially ejected through the exit slits in the X-rods. The supplementary ac signal also contributes to the ion resonance excitation.

Characteristics of the LTQ

The LTQ offers much better performance over the conventional 3D ion trap¹²⁰: 15 times higher ion capacity, 3 times faster scan rate, up to 100% detection efficiency, and up to 70% trapping efficiency. These tremendous advances can be directly translated into improved identification results in proteomics study^{123, 124}.

Performance of the LTQ was examined by Gorshkov *et al*¹²⁵. According to their and our experience, the LTQ is able to achieve a mass accuracy of up to 50 ppm and resolution close to 1000 (m/z 400), at an ion target value of 5,000 (number of ions filling the trap). In general, the LTQ is close to ideal for proteomics studies due to its fast scan rate and high sensitivity. However, major caveats of ion traps – when used alone – include low resolution, low mass cutoff, and modest mass accuracy.

1.2.2.3 Fourier transform ion cyclotron resonance

Principles of the the FTICR

Fourier transform ion cyclotron resonance (FTICR) measures ion oscillation frequencies in a combined magnetic field and electric field. Superconducting magnets with field strength of 3 Tesla or higher are able to provide a uniform, unidirectional and homogeneous magnetic field over time. The cubic analyzer cell is composed of six plates (Figure 1.10). Two trapping plates are perpendicular to the magnetic field. Two ion excitation plates and two detection plates are positioned in parallel to the magnetic field. An ultra-high vacuum of 10^{-9} to 10^{-10} mBar is required for ion detection.

The force that an ions experience in a combined magnetic and electric field is:

$$(4) \quad F = qE + q(v \otimes B)$$

Where q is the ion charge; E is the electric field strength; v is the ion velocity; B is the magnetic field strength.

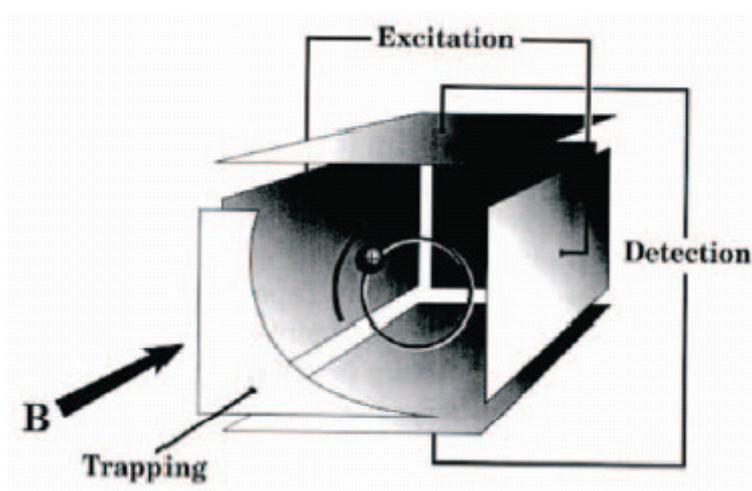


Figure 1.10 Structure of the FTICR analyzer cubic cell. B: magnetic field.

Cyclotron motion

An ion entering the cubic cell experiences zero force in the Z direction, i.e. direction of the magnetic field B. However in the direction perpendicular to the magnetic field B, the Lorentz force causes the ion to periodically oscillate in circles. The frequency of this motion, termed cyclotron motion, is independent of the ion's kinetic energy but depends on the m/z value of the ion:

$$(5) \quad f_c = \frac{q \cdot B}{2\pi \cdot m}$$

Where m is the ion mass; q is the ion charge; B is the magnetic field.

Kinetic energy influences the radius of the cyclotron orbit. The radius scales with the square root of the ion's kinetic energy^{110, 111}.

Trapping motion

A small (~0.2 V), symmetric positive voltage is applied to the trapping plates to trap positive ions. This weak electrical field generates a negative gradient starting from the trapping plates towards the center of the cubic cell. Ions oscillate between trapping plates in simple harmonic motion, with the trapping motion frequency of

$$(6) \quad f_T = \sqrt{\frac{q \cdot \alpha \cdot V_T}{\pi^2 \cdot m \cdot a^2}}$$

Where m is the ion mass; z is the ion charge; V_T is the small symmetric voltage; α and are the constants representing the ICR cell geometry.

Magnetron motion

The electric potential from the two trapping plates distributes among the six plates of the ICR cell and generates an uneven potential distribution in space. At the XY-plane, this potential decreases when it goes away from the center of the cell. Therefore ions tend to drift away from the center with the frequency of

$$(7) \quad f_m = \frac{\alpha \cdot V_T}{\pi \cdot a^2 \cdot B}$$

Where V_T is the small symmetric voltage; α and are the constants representing the ICR cell geometry; B is the magnetic field.

Due to the magnetron motion, the actual detected frequency is the difference between cyclotron frequency and magnetron frequency. However, the magnetron frequencies are in the order of 1-100 Hz, whereas cyclotron frequencies are of 5 kHz – 5 MHz. Therefore magnetron frequencies impose only minute influence on the cyclotron frequencies. As seen from formula (7), magnetron frequency is independent of the m/z value, but the radius of ion magnetron motion depends on the ion's initial position within the ICR cell¹¹⁰.

Ion excitation and detection

For detection purpose, the radius of the ion cyclotron motion has to be as large as possible. A sinusoidal voltage is applied to the excitation plates. If the frequency of ions ($f_c - f_m$) matches the applied rf frequency, the ions will absorb energy and oscillate in resonance. They spiral with increasing radius towards the detection plates. After the excitation rf is removed, ions cycle in a stable manner and are ready for detection. Their periodic motion produces a sinusoidal image signal in the detection plates that can be amplified, recorded and processed by Fourier transform algorithm to generate mass spectra.

Characteristics of FTICR

Characteristics of FTICR instruments are listed in Table 1.2. In general, the parts per million (ppm) mass accuracy, high resolution, and high dynamic range are the great advantages of FTICR. A resolution of 10^6 was achieved for ions with m/z larger than 1000, by scanning with longer transients (>60 seconds) under high vacuum of 10^{-10} mBar¹²⁶. However, the drawbacks of FTICR have been slow scanning rate, relatively low sensitivity, and space-charge effect. Recently larger cubic cells have become commercially available for reduced space-charge effect and improved dynamic ranges.

Table 1.2 Characteristics of FTICR MS

| Mass accuracy | Resolution @ 1 s scans | Sensitivity | Dynamic range |
|---------------|------------------------|-------------|---------------|
| 1ppm | 100,000 | 0.5 fmol | 5000 |

1.2.2.4 Orbitrap

Principles of the Orbitrap

The Orbitrap is a new type of mass analyzer which resembles the conventional Kingdon trap, already known since 1924, to some degree, but embraces a new concept of trapping and analyzing ions.

The Orbitrap is composed of a spindle-shaped inner electrode and a split, barrel-like outer electrode (Figure 1.11). The largest inner diameters for the inner electrode and for the inner surface of the outer electrode are 8 mm and 20 mm, respectively¹²⁷. A static, quadrupole electrostatic field is generated by the Orbitrap's axially symmetric electrodes:

$$(8) \quad U(r, z) = \frac{k}{2} \left(z^2 - \frac{r^2}{2} \right) + \frac{k}{2} (R_m)^2 \ln \left[\frac{r}{R_m} \right] + C$$

Where r and z are cylindrical coordinates ($z = 0$ being the plane of the symmetry of the field); C is a constant; k is the field curvature; R_m is the characteristic radius.

Ions are pulsed injected into the Orbitrap from an entrance slit offset from its equator. Without the need for additional excitation, ions start to oscillate with the combined motions along the axial direction (Z direction) and in rotational plane (perpendicular to the Z axis). To squeeze ions into tight ion packets and to move them far enough from the out electrodes, an initial adjustment by increasing electric fields is performed, typically lasting for 20-100 μs ^{112, 127, 128}.

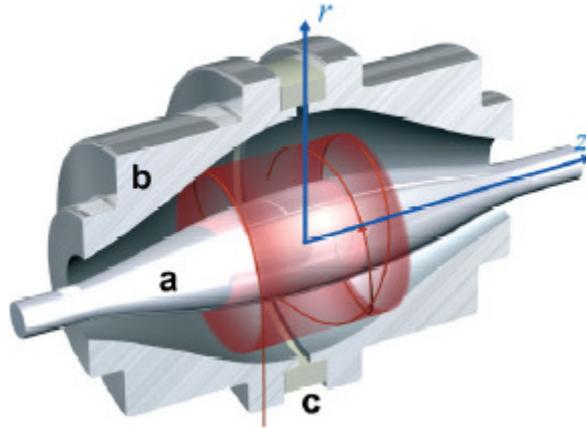


Figure 1.11 Cut-away structure of the Orbitrap mass analyzer. From Scigelova *et al.*¹²⁹.

Rotational motion and ion trapping

Due to the electric field, ions experience centrifugal and centripetal forces in the plane perpendicular to the Z direction. This rotational motion defines a frequency of

$$(9) \quad \omega_{\phi} = \omega \sqrt{\frac{\left(\frac{R_m}{R}\right)^2 - 1}{2}}$$

Where R_m is the characteristic radius; R is the radius of the circle; ω is the frequency of axial oscillations.

And the radius of the rotation is

$$(10) \quad r = 2 \frac{eV}{eE}$$

Where eV is the ion's kinetic energy; eE is the force due to the electric field which is directly radially inward.

Equation (10) indicates that, once the field strength is fixed for a certain electric sector, the radius of the ion rotational motion depends on the ion kinetic energy. In other words, only certain ions with appropriate kinetic energies obtain stable trajectories along the rotational plane.

Radial ion motion

The radial component of the electric field also influences the ion's orbital trajectory. The frequency of the radial motion is described by equation (11). The initial position, as well as energy distributions in the rotational and radial directions can directly influence the radial motion¹¹². Therefore with time the ion radial motion will go out of phase, with a speed of orders of magnitude faster than in the axial direction.

$$(11) \quad \omega_r = \omega \sqrt{\left(\frac{R_m}{R}\right)^2 - 2}$$

Where R_m is the characteristic radius; R is the radius of the circle.

Axial ion motion and ion detection

Tight ion clouds can oscillate along the Z direction in-phase, with the frequency described in equation (12). This axial frequency is independent of kinetic energy, initial position and other ion parameters. Therefore it is unique to each given m/z and serves the purpose of ion detection and mass measurement¹¹².

$$(12) \quad \omega = \sqrt{(q/m)k}$$

Where m is the ion mass; q is the ion charge; k is the field curvature.

Due to the symmetric shape of the electrodes, the image currents generated by the dephased rotational and radial oscillations in different sections will eventually cancel each other out. However, the axial oscillation persists to generate useful signal current, which can be translated to mass spectra using the Fourier transform algorithm.

Characteristics of the Orbitrap

Characteristics of Orbitrap are listed in Table 1.3. The Orbitrap ‘inherits’ all the advantages of FTICR, i.e. high mass accuracy, high resolution, and high dynamic range while being a much smaller instrument and not needing a superconducting magnet. Due to the unique structure of the LTQ-Orbitrap instrument, an internal calibration method using a lock-mass can achieve sub-ppm mass accuracy¹³⁰. Compared to the FTICR, the Orbitrap has increased sensitivity and much reduced space-charge effects. The latter is due the fact that ions with different kinetic energies generate different rotational motions and ions are shielded from each other by the central electrode. Ion thin rings with different m/z will disperse out, limiting the columbic repulsion effects.

Table 1.3 Characteristics of the Orbitrap MS.

| Mass accuracy | Resolution @ 1 s scans | Sensitivity | Dynamic range |
|---------------|------------------------|-------------|---------------|
| 1ppm | 60, 000 | 0.1 fmol | 5000 |

1.2.2.5 Hybrid mass spectrometers to study peptide primary sequences

Hybrid mass spectrometers are commonly used in proteomics studies. The LTQ-FT and LTQ-Orbitrap combine the advantages of fast scanning rate and high sensitivity of the LTQ and the advantages of high mass accuracy, high resolution, and high dynamic range of the FTICR and Orbitrap^{131, 132}. To reduce space-charge effect, automatic gain control (AGC), which delivers a constant number of ions per scan, is used in the LTQ before injection into the FTICR or Orbitrap. This has been demonstrated to produce highly accurate m/z measurements by cancelling space charge effects. The hybrid nature of the instrument also facilitates fast peptide fragmentation in space.

Peptide fragmentation generates very useful information of amino acid sequences. Collision induced fragmentation (CID) is the most widely used method. CID introduces low energy fragmentation via collision with inert gases (e.g. helium) at energies of some eV. Repeated collisions cause peptides to decompose. Breakage of peptides can happen in several modes

(Figure 1.12). In our electrospray-LTQ-FT/Orbitrap setup of analyzing tryptic peptides, CID mostly generates b and y ions.

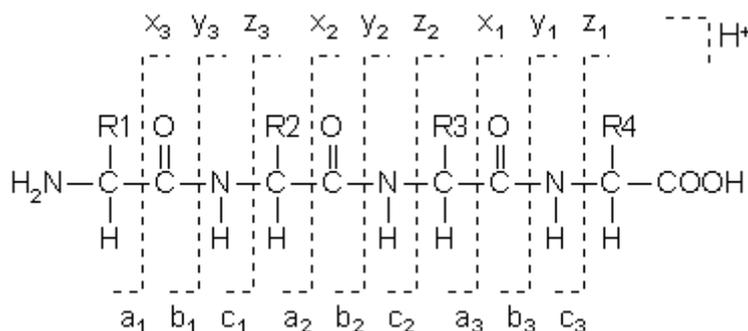


Figure 1.12 Nomenclature for fragment ions^{133, 134}. Peptide back bones are shown with amino acid side chains represented with R. CID usually generates y and b ions.

An ion detection cycle in our MS-analysis is composed of two parts, the full scan and the peptide fragmentation scan. Precursor ions are scanned in the FTICR or Orbitrap for the entire range of m/z values (MS). Ions of interest are then separately isolated and accumulated in the linear ion trap for CID peptide fragmentation (MS/MS). By retaining fragment ions of interest in the trap via ion selective instability mode, further fragmentation (MS^n) can in principle also be performed. This has been proven very useful in deriving sequence information from phosphorylated peptides¹³⁵. Recently, multistage activation (MSA) has become available. In this special scan mode, the MS/MS/MS fragmentation is generated in the presence of all the MS/MS fragments without intervening isolation step. Therefore the resulting mass spectra are a combination of MS^2 and MS^3 fragments. MSA is routinely used in our phosphorylation studies. Due to the fast scan rate in LTQ, the MS event in FT/Orbitrap and MS^n events in LTQ can be easily combined without extending the cycle duty¹³⁵.

Other types of fragmentation techniques are also used, such as electron capture dissociation (ECD) and electron transfer dissociation (ETD). Electron transfer generates a completely different pathway of thermo-energy distribution and peptides tend to fragment at the $\text{C}_\alpha\text{-NH}$ bond. ECD and ETD are suggested to generate complementary mass spectra to those from CID¹³⁶.

1.2.3 Quantitative proteomics

Biological research often requires the knowledge of protein amounts and their changes under different conditions. To meet these needs, various quantitation methods have been developed in the past few years. MS-quantitation that entirely depends on the signal intensity of unlabeled peptides, i.e. label-free quantitation, is just emerging. This method demands a very precise and accurate performance of the whole proteomics workflow. In contrast, a large group of other methods employ stable isotopes for labeling peptides. They usually better tolerate the signal fluctuations in HPLC and mass spectrometers.

Stable isotope encoding changes the physiochemical properties of the peptides by the least possible amount. Protein/peptide samples that are labeled with stable isotopes have shifted m/z values when compared to their natural, non-isotope-labeled counterparts but are otherwise identical in all respects. Thus, stable isotopes such as ^{13}C , ^{15}N , and ^{18}O do not induce shifts in HPLC retention times¹³⁷. Therefore labeled and non-labeled peptides show up as pairs in mass spectra. Their relative intensities can be directly visualized¹³⁸. Deuterated peptides shift slightly from their non-deuterated counterparts¹³⁹. However, this problem can be corrected if quantitation is based on the entire elution profiles of HPLC instead of on a few single observations¹⁴⁰. Generally, there are three ways to label proteins or peptides with stable isotopes (Figure 1.13)^{141, 142}. Metabolic labeling supplies stable isotopes during the growth and development of cells^{138, 143} and organisms¹⁴⁴⁻¹⁴⁶. Chemical labeling modifies certain amino acid side chains with natural or isotope-labeled reagents^{147, 148}. Enzymatic labeling uses trypsin or Glu-C catalyzed incorporation of ^{18}O during protein digestion^{149, 150}. To distinguish the labeled and non-labeled forms of peptides by MS, it is recommended to generate a minimum of 4 Da difference in the peptide pairs.

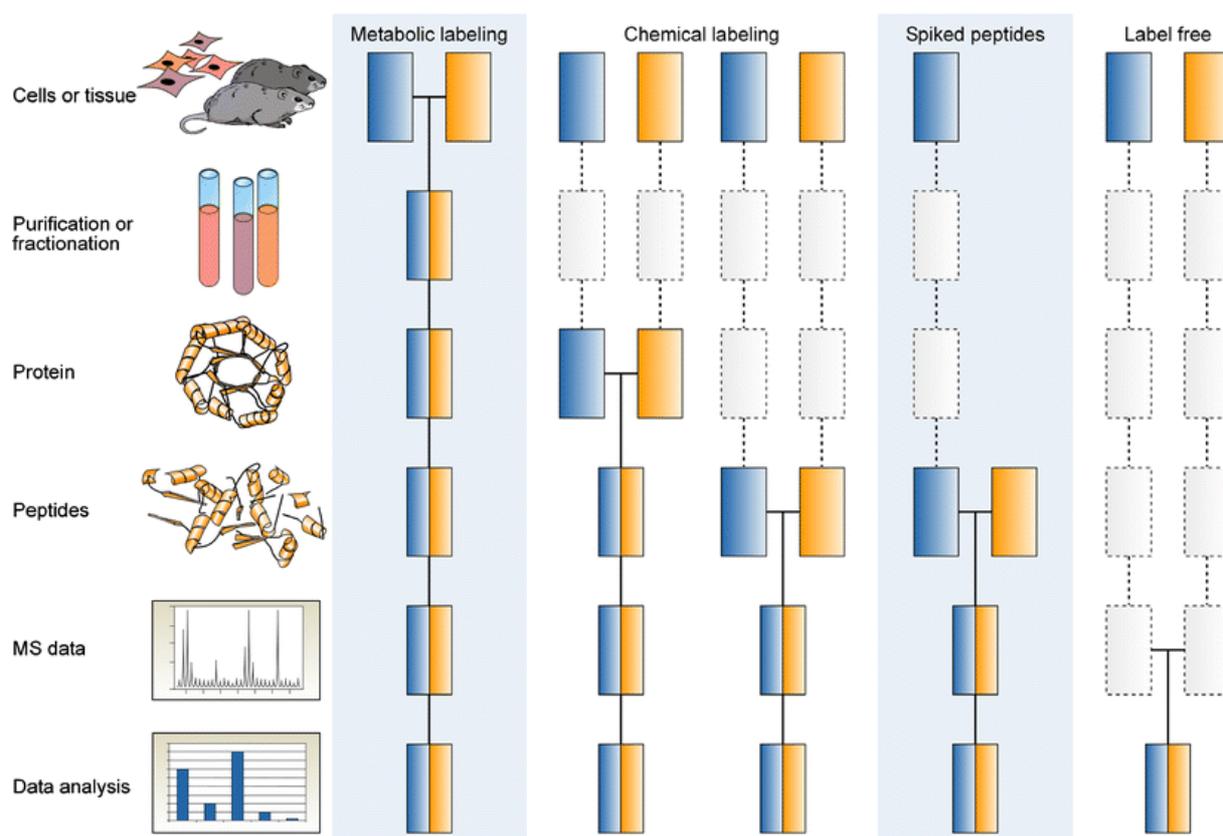


Figure 1.13 Strategies in mass spectrometry-based quantitative proteomics. Boxes in blue and yellow represent two experimental conditions. Horizontal lines indicate the experimental step where samples are combined. Dashed lines indicate points at which experimental variation and thus quantification errors can occur. From Bantscheff *et al.*¹⁴¹.

It is advantageous to mix samples at an early stage of the experiment to avoid accumulating systematic errors, which can translate into inaccurate quantitation results in the sensitive mass spectrometry measurement. From this perspective, metabolic labeling is superior to other quantitation methods. Stable isotope labeling with amino acids in cell culture (SILAC) is an established method with all the strengths of the metabolic labeling strategy^{138, 151} (Figure 1.14). SILAC labeling utilizes arginine and lysine with heavy elements of ^{13}C , ^{15}N , and ^2H . The most commonly used forms are $^{13}\text{C}_6\text{-Arg}$, $^{13}\text{C}_6^{15}\text{N}_4\text{-Arg}$, $^2\text{H}_4\text{-Lys}$ and $^{13}\text{C}_6^{15}\text{N}_2\text{-Lys}$. Up to three different biological conditions can be compared in a single SILAC experiment and many interesting biological discoveries have been obtained using this method¹⁵¹.

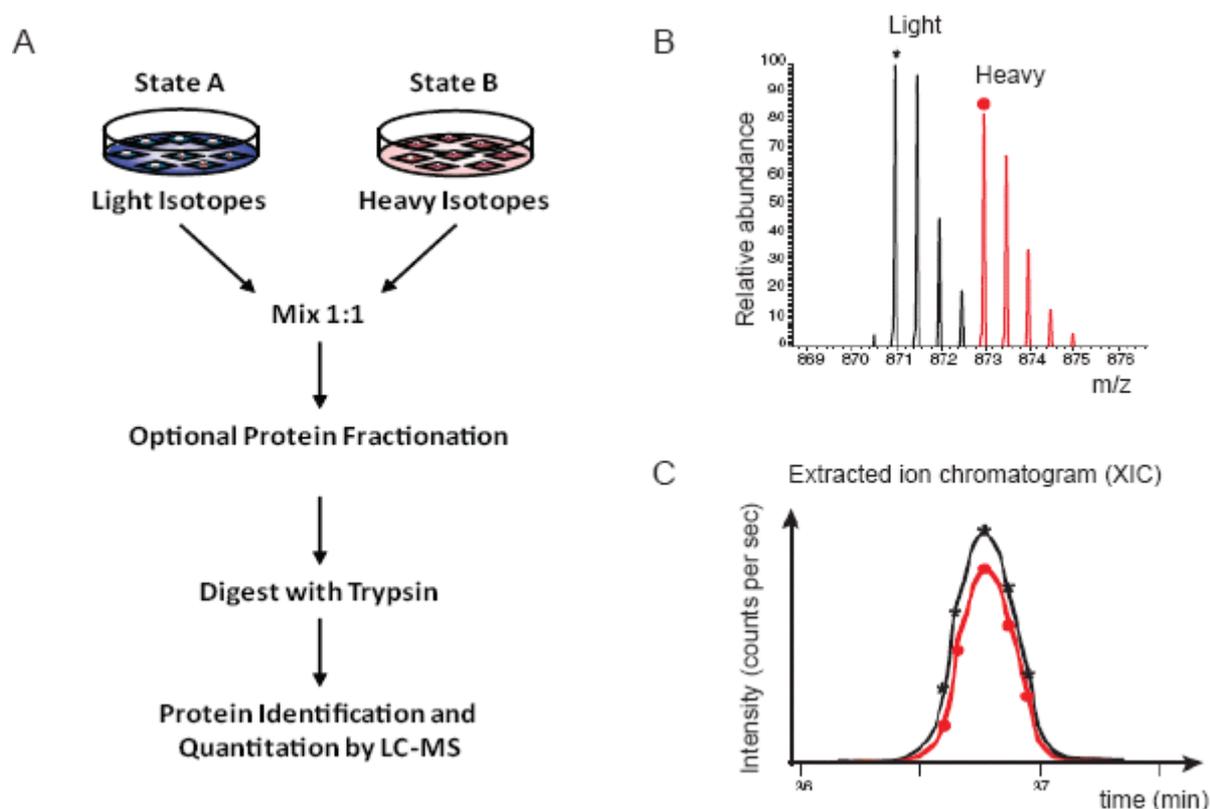


Figure 1.14 SILAC strategy in quantitative proteomics. (A) Two SILAC states are presented. One state is labeled with light isotopes (e.g. $^{12}\text{C}_6^{14}\text{N}_4$ -Arg and/or $^{12}\text{C}_6^{14}\text{N}_2$ -Lys) and the other state with heavy isotopes (e.g. $^{13}\text{C}_6^{15}\text{N}_4$ -Arg and/or $^{13}\text{C}_6^{15}\text{N}_2$ -Lys). Samples are mixed as early as possible to avoid introducing experimental errors. (B) Relative quantities of the peptides are visualized in MS spectra. (C) Extracted ion chromatogram indicates that light- and heavy-labeled peptides co-elute very well during reverse phase HPLC separation. Modified from Ong *et al.*¹⁴².

iTRAQ (an isobaric tag for relative and absolute quantitation) is a chemical labeling method which can quantify up to eight different conditions in single experiment¹⁵². A multiplexed set of isobaric reagents are used to modify each sample. Differently modified peptides display no difference in MS scans due to the isobaric property of the mass tags. However, after peptide fragmentation the unique, low-mass reporter ion of each tag (113-121 Da) is displayed in MS/MS spectra. Quantitation is therefore carried out in MS/MS spectra rather than MS spectra.

Absolute quantitation is another interesting challenge in quantitative proteomics. Using a reference sample with known amount, the aforementioned quantitation methods can be used

to deduce the absolute amount of proteins in the sample. Successful examples include the precise quantitation of Grb2 copy numbers in HeLa cells by means of absolute-SILAC¹⁵³.

1.2.4 Phosphoproteomics

Phosphorylation is thought to occur in one third of the proteome. Early estimates suggest that there are over 100,000 phosphorylation sites in human proteome. While ubiquitously distributed, phosphoproteins are typically of low abundance. Furthermore, often only a few percent of the entire protein amount is present in a phosphorylated form¹⁵⁴. Phosphorylation governs most signal transduction processes and many other cellular functions. Therefore the study of phosphoproteomics is of great importance.

To enrich the low abundant phosphorylated proteins or peptides, various methods have been developed. Chromatographic methods include affinity binding to phospho-antibodies^{27, 155}, affinity binding to kinase domains¹⁵⁶, metal chelation^{94, 157, 158}, and ion exchange¹⁵⁹. Chemical modifications of the phosphate group are also used to create affinity tags for purification¹⁶⁰⁻¹⁶³. Engineering proteolytic digestion at the phosphorylation sites is also proposed^{164, 165}. Among these, the chromatography methods are most widely adopted due to their high enrichment efficiency and experimental simplicity.

1.2.4.1 Antibody-based enrichment

Tyrosine phosphorylation comprises less than 2% of the cellular phosphorylation events³¹. Fortunately good quality antibodies of phospho-tyrosine (pTyr), such as 4G10 and pTyr100, can be employed to enrich pTyr containing proteins or peptides^{27, 93, 155}. Antibodies generated against kinase substrate motifs, such as AKT motif, are also used to enrich substrates of specific kinases¹⁵⁶.

1.2.4.2 Strong cation exchange (SCX)

SCX is a low resolution but robust enrichment method^{159, 166}. The principle of using SCX in phosphopeptide analysis is based on reduced positive charges on the phosphorylated peptides. Most tryptic peptides carry one positive charge at each peptide terminus at pH 2.7, as specified in the SCX buffer (NH_4^+ from the N-terminal amino group and the positively

charged side chain of trypsin or lysine). The negatively charged phosphate group can counter the positive charges, effectively reducing the charge state by one, and therefore decrease the binding to the SCX column. Generally multiply phosphorylated peptides bind to the column with minimum affinity, while non-phosphorylated peptides bind strongly. However, acidic amino acids (glutamic acid and aspartic acid) can interfere with this strategy. Gygi and coworkers demonstrated large scale identification of 2,001 phosphopeptides using SCX fractionation¹⁶⁶.

1.2.4.3 TiO₂ enrichment

Titanium dioxide (TiO₂) particles are stable with regards to mechanical, chemical and thermal stress. Their unique amphoteric ion-exchange properties suggested their use as an alternative material to silica in chromatography¹⁶⁷. In recent years, its specific affinity to organic phosphates in solution has been revisited. Heck and coworkers demonstrated that TiO₂ chromatography can achieve very high enrichment efficiency (90%) for phosphopeptides in simple samples¹⁶⁸. For complex samples, however, non-specific binding of the acidic amino acid residues such as glutamic acid and aspartic acid becomes significant. Larsen et al. proposed the use of 2,5-dihydroxy benzoic acid (DHB) to compete with the acidic peptides from binding⁹⁶. Because the binding strengths decrease from phosphopeptide to DHB to acidic peptides (Figure 1.15), this approach has proven to be very successful in large scale phosphoproteomic studies³¹.

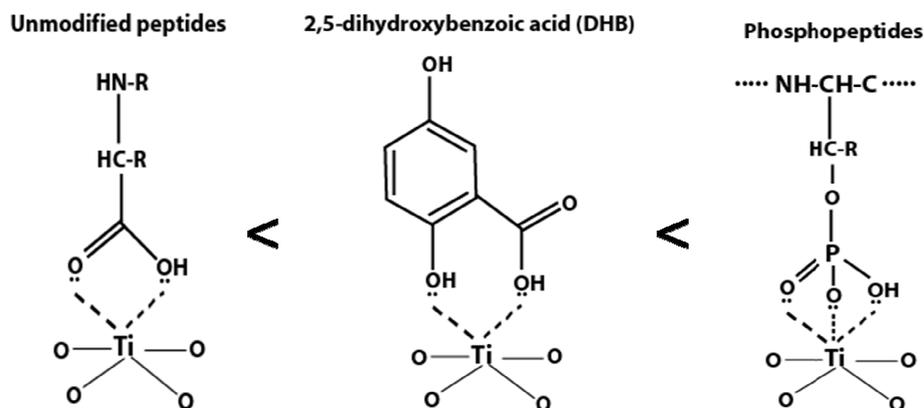


Figure 1.15 Comparison of binding capacities to titanium dioxide sphere. The binding capacities increase from carboxylic group, DHB, to phosphate group. Adapted from *Olsen JV (ASMS 2005)*.

1.2.4.4 Phosphorylation beyond Ser, Thr, and Tyr

In cells, phosphate groups are predominantly attached to the hydroxyl groups in serine, threonine and tyrosine residues. These are termed the O-phosphorylation. There are also a variety of other phosphorylation modes in cells¹⁵⁴. In N-phosphorylation, phosphate groups attach to histidine and lysine. S-phosphorylation occurs on cysteine. Acyl-phosphotates link to aspartic and glutamic acid. New types of phosphorylation have been reported, such as N- or O- phosphorylation on hydroxyl-lysine.

In our phosphoproteomics workflow, acid conditions are employed. This can lead to general hydroxylation of most uncommon phosphorylations, e.g. histidine phosphorylation¹⁶⁹. To investigate those phosphorylation events, basic buffer conditions plus negative ion mode are recommended.

1.2.4.5 Quantitative phosphoproteomics

The recent years have witnessed a breakthrough in phosphoproteomics, such that hundreds or thousands of phosphorylation sites can be obtained in single experiments^{31, 94, 96, 159, 166}. Most of these studies contain qualitative information of phosphorylation site identifications rather than quantitative information.

The first large scale quantitative, site specific and time-resolved phosphoproteomics study was reported by Olsen *et.al*, where the phosphoproteome changes after EGF stimulation were quantified at several time points³¹. In that study, phosphorylation distribution among Ser, Thr and Tyr residues was measured. Dual phosphorylation in kinase activation loops were captured, which directly visualized the activation state of the enzymes. Time-resolved studies revealed trafficking of phosphoproteins between cytoplasm and nucleus (e.g. STAT5 and MAPK1/3). Clustering of phosphopeptides based on their reaction kinetics indicated functional groups. Rich biological information was mined from the long list of phosphorylated proteins and sites. This work clearly demonstrated the strength of quantitative phosphoproteomics.

Based on the work of Olsen *et al*, we established a general workflow of quantitative phosphoproteomics, which is depicted in Figure 1.16.

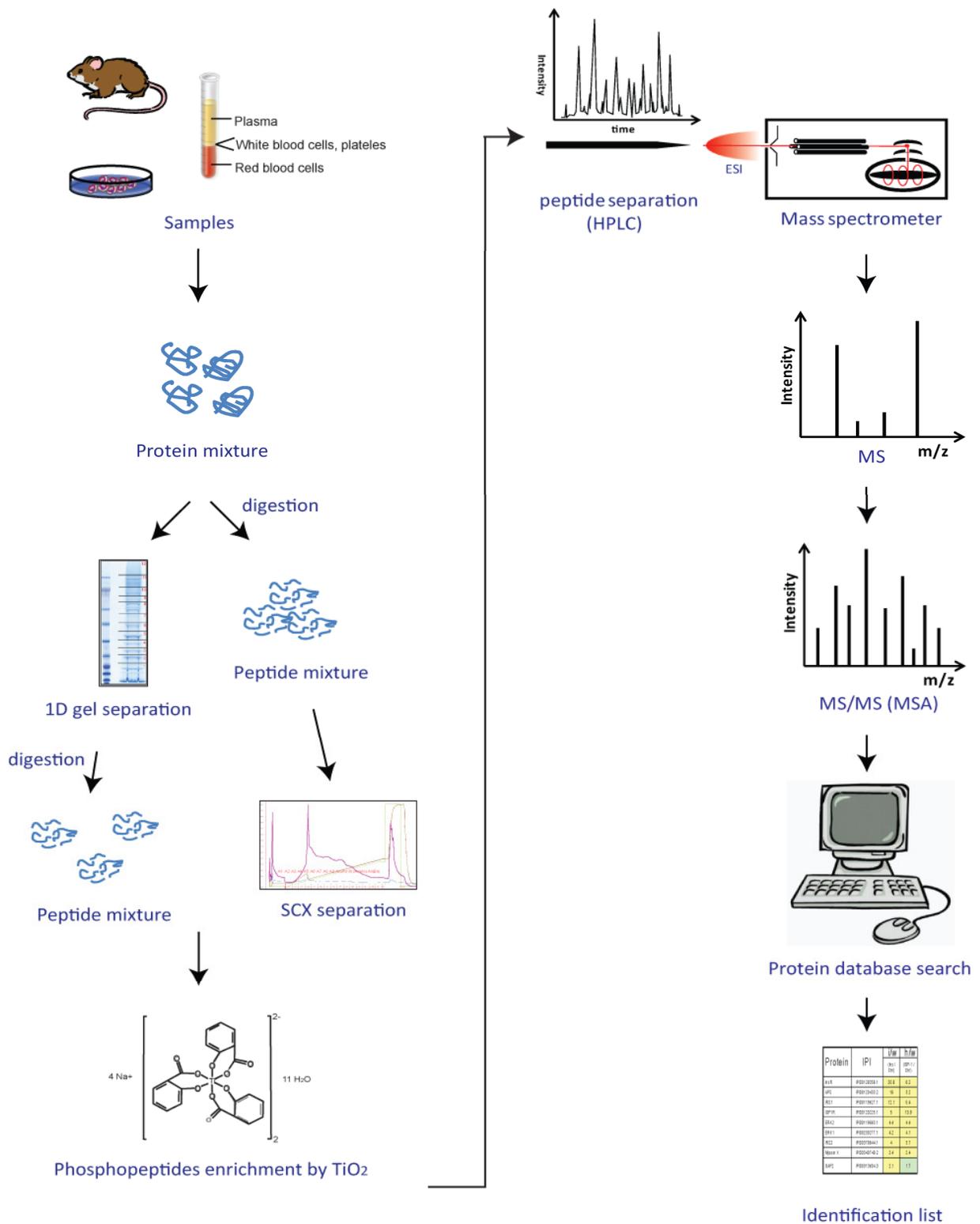


Figure 1.16 The general workflow of mass spectrometry-based phosphoproteomics.

2 Aim of the Study

2.1 Cancer cell line to evaluate inhibitors for kinases and phosphatases

SILAC-based quantitative phosphoproteomics was previously demonstrated to be a powerful technique to decipher phosphorylation changes after growth factor stimulation in cell lines³¹. In principle, essentially the same technique can be applied to very practical topics such as the effect of inhibitor compounds used in everyday research and clinics.

In this thesis, project 1 is designed as a proof-of-principle study to quantify the effects of several commonly used phosphatase inhibitors in laboratories. Meanwhile it is a technology-centered study, with the aim to analyze the phosphoproteome in depth. Project 2 is designed to tackle the effects of some described kinase inhibitors on the entire signaling network, including one clinical drug.

2.2 Liver cell model to assess normal and transformed cells

Elucidation of the properties of transformed cells is a long standing topic in cancer research. Some of these cell lines were derived from tumor tissues, and therefore should retain most properties of the tumors. Project 3 is designed to characterize the phenotypes of transformed cancer cell line with comparison to primary, non-transformed cells.

3 Materials and Methods

3.1 Cell models

3.1.1 Immortalized cell lines

Mouse hepatoma cell line Hepa1-6

Hepa1-6 was derived from the transplantable hepatoma BW7756, which was chemically introduced in C57LBL/6 mice. These cells are adherent and grow as monolayer in culture. Hepa1-6 is thought to maintain the major properties and functions of liver, such as secreting several liver-specific products including albumin, alpha1-antitrypsin, alpha-fetoprotein, and amylase¹⁷⁰. As demonstrated by several studies, Hepa1-6 has poor immunogenicity^{171, 172}. It is highly sensitive to adenovirus infection and the antisense treatment targeting the oncogene IGF-1. It is highly sensitive to adenovirus infection and susceptible to viral replication, progeny production and cytopathic effect. Therefore it has become a popular tumor model for experimental immunotherapy studies.

Human epithelial carcinoma cell line HeLa

HeLa is descended from cervical carcinoma transformed by human papillomavirus 18 (HPV18) in a female patient. Rearrangement of chromosomes was observed. They have a modal chromosome number of 82, with four copies of chromosome 12 and three copies of chromosomes 6, 8, and 17. HeLa cells are adherent and maintain contact inhibition *in vitro*.

Human myelogenous leukemia cell line K562

K562 was the first human immortalized myelogenous leukemia line to be established. It originated from a female CML patient in blast crisis phase and contains the e14a2 fusion protein BCR-ABL. This fusion protein is the product of the Philadelphia chromosome which is generated by fusing the long arms of chromosome 9 and 22. These two chromosomes encode ABL and BCR, respectively. BCR-ABL exhibits constitutive tyrosine kinase activity

of ABL and has been proven to be essential in CML. Therefore K562 cells are tumor models for experimental CML studies.

K562 cells exhibit a second reciprocal translocation between chromosome 15 and chromosome 17. They resemble both undifferentiated granulocytes¹⁷³ and erythrocytes¹⁷⁴. They are deficient in major histocompatibility complex (MHC) and they lack herpesviruses. K562 cells are rounded and grow in suspension.

3.1.2 Primary hepatocytes

Isolation and culture of mouse hepatocytes were performed according to standard operation procedures established in the German HepatoSys Network and briefly summarized here¹⁷⁵. Briefly, six to twelve weeks old B6 mice were used. The use of mice for hepatocyte isolation has been approved by the animal experimental committees and animals were handled and housed according to specific pathogen free conditions. Anesthesia was achieved by intraperitoneal injection of 5mg/100mg body weight ketamine hydrochloride 10% (115.34 mg/ml; Essex Tierarznei, Munich, Germany) and 1mg/100mg body weight xylazine hydrochloride 2% (23.32 mg/ml; Bayer Leverkusen). HANKS solution I was produced by supplementation of BASAL HANKS solution (8 g NaCl, 0.4 g KCl, 3.57 g Hepes, 0.06 g Na₂HPO₄ x 2 H₂O, 0.06 g KH₂PO₄ in 1 L distilled H₂O, adjusted pH to 7.4, sterilized) with 2.5 mM EGTA, 0.1% glucose and penicillin/streptomycin at a dilution of 1 : 100. HANKS solution II was produced by supplementation of BASAL HANKS solution with 0.3 mg/ml collagenase CLSII and 5 mM CaCl₂. Hanks solutions I and II were prewarmed in a 42°C waterbath. Collagenase was added immediately prior to liver perfusion. After shaving and cleaning the abdomen under sterile conditions, the abdominal cavity was opened and the portal vein was cannulated with a 24G catheter. A silicon tube (diameter 2.4 mm) was connected to the catheter and HANKS solution I was infused via a peristaltic pump at a flow rate of 8 ml per minute. After starting the peristaltic pump, the vena cava and the right heart ventricle were incised to permit sufficient outflow. The liver was perfused with solution I for 5 minutes, followed by HANKS solution II for 5-7 minutes. Correct placement of the portal vein catheter is evidenced during perfusion by the steady and even change from dark red-brown to a light brown color in all liver lobes. Following perfusion the liver was transferred to a sterile Petri dish and the gall bladder was removed. The following steps were performed in a sterile hood. The liver capsule was carefully removed using a pincette. Gentle shaking

disintegrated the perfused liver, yielding a suspension of single cells, some cell clumps and cell debris. The suspension was placed onto a 100 μm cell strainer and was filtered through the mesh by gravity flow. The suspension was transferred to a 50 ml Falcon tube and washed twice with Williams E medium (room temperature). Cells were centrifuged at room temperature at $37.5 \times g$ for 2 minutes in a cell culture centrifuge. The cells were resuspended in Williams medium E and the percentage of intact cells was determined by staining an aliquot with trypan blue. On average 70-80% of the cells were viable.

Hepatocytes were placed on collagen coated tissue culture dishes in *FCS cell culture medium* (William's medium E supplemented with 10 % FCS, 100 nM dexamethasone, 2 mM L-glutamine and 1% penicillin/streptomycin solution) and kept in a humidified cell culture incubator at 37°C and 5% CO₂ (Figure 1). For 6 cm (diameter) tissue culture dishes hepatocytes were plated at a density of 2×10^6 cells/dish in 3 ml FCS-culture-medium. After 4 h of incubation hepatocytes should be attached to the collagen coated dish. Subsequently, the FCS cell culture medium was removed and replaced by *serum free cell culture medium* (William's medium E supplemented with 100 nM dexamethasone, 2 mM L-glutamine and 1% penicillin/streptomycin solution).

In project 3, after cultivation for 14 hours, the primary hepatocytes were placed on ice and the medium was removed. The cells were lysed in modified RIPA buffer.

Materials for isolation and culture of primary hepatocytes: Ultra pure bovine serum albumin (BSA), dexamethasone and insulin were obtained from Sigma (Deisenhofen, Germany). Fetal calf serum (FCS), penicillin/streptomycin (10.000 U/ml - 10mg/ml) and L-Glutamin from Gibco (Paisley, Scotland) and William's medium E (WME) from Biochrom (Berlin, Germany). Materials used for hepatocyte isolation and culture were Abbocath-T 18G (Venisystems, Abbott, Ireland), Peristaltic pumps (Reglo Digital; Isamtech, Zürich, Switzerland), 100 μm meshcell strainers (Becton-Dickinson, Heidelberg), collagen I coated 6-well-plates (Becton-Dickinson, Heidelberg) and tissue culture centrifuges (Labofuge M, Heraeus, Stuttgart, Germany). All other chemicals of analytical grade were obtained from Merck (Darmstadt, Germany).

3.1.3 Cell culture

Hepa1-6 and HeLa cells were obtained from the American Type Culture Collection (ATCC). K562 cells were kindly provided by Henrik Daub. General cell culture conditions were 37 °C, 5% CO₂, and humidified atmosphere.

Hepa1-6 and HeLa cells were cultured in DMEM medium (4.5 g/l glucose) supplemented with 10% fetal bovine serum, 1% (10 mg/ml) streptomycin / (10,000 U/ml) penicillin, 1% L-glutamine, and 4.5 g/l glucose. Cells were seeded with 20-30% density and were split or harvested when the cell density reached around 90%.

K562 cells were cultured in suspension in RPMI medium supplemented with 10% fetal bovine serum, 1% (10 mg/ml) streptomycin / (10,000 U/ml) penicillin, 1% L-glutamine, and 4.5 g/l glucose. Cells were seeded with the density of 0.2×10^6 cells/ml, and were split or harvested at the density of $1-1.2 \times 10^6$ cells/ml.

The following forms of stable isotope-labelled arginine and lysine were used: L-¹³C₆-arginine (Arg6), L-¹³C₆¹⁵N₄-arginine (Arg10), L-²H₄-lysine (Lys4), and L-¹³C₆¹⁵N₂-lysine (Lys8). Normal arginine and lysine used were: L-¹²C₆¹⁴N₄-arginine (Arg0) and L-¹²C₆¹⁴N₂-lysine (Lys0). To generate double encoding SILAC conditions, normal medium deficient in arginine and lysine was supplemented with Arg10 and Lys8 for the “heavy” condition, or with Arg6 and lys4 for the “medium” condition, or with Arg0 and lys0 for the “light” condition. Final concentrations of arginine are 28 mg/l in DMEM and 84 mg/l in RPMI. Final concentrations of lysine were 73 mg/l in DMEM and 49 mg/l in RPMI. In each SILAC condition, medium was supplemented with 10% dialyzed fetal bovine serum with 10 kDa cutoff, 1% (10 mg/ml) streptomycin / (10,000 U/ml) penicillin, and 1% L-glutamine.

SILAC cell culture was essentially the same as described for normal cell culture, except that SILAC medium and dialyzed FBS were used. Generally, more than 95% incorporation of the labelled amino acids can be achieved after 6 passages of cell proliferation.

Materials for cell culture: DMEM and RPMI medium were obtained from Gibco BRL. Medium without Arg and Lys was obtained from PAA Laboratories and SAFC Biosciences. L-arginine, L-lysine, L-¹³C₆¹⁵N₄-arginine and L-¹³C₆¹⁵N₂-lysine were purchased from Sigma-Aldrich. Fetal bovine serum and dialyzed fetal bovine serum dialyzed were obtained from

Invitrogen. L-glutamine (200 mM in 0.85% NaCl) and streptomycin (10 mg/ml)/penicillin (10,000 U/ml) were products of Invitrogen.

3.1.4 Preparation of inhibitor compounds

Calyculin A, deltamethrin, and Na-pervanadate were prepared freshly before the experiment. To prepare stock solutions, calyculin A was dissolved in DMSO to 100 μ M and deltamethrin was dissolved in DMSO to 10 mM. In project 1, these two phosphatase inhibitors were applied to cells with final concentrations of 100 nM for calyculin A and 10 μ M for deltamethrin. To prepare Na-pervanadate, 30% H₂O₂ was diluted in 20 mM HEPES, pH 7.3 to 100 mM, then mixed with an equal volume of 100 mM sodium orthovanadate. After five minutes, a scoop of catalase was added to the pervanadate stock to neutralize extra H₂O₂. In project 1, Na-pervanadate was applied to cells to a final concentration of 0.5 mM.

The 10 mM stock solutions of U0126 and SB202190 were prepared by dissolving the chemicals in DMSO. In project 2, these two kinase inhibitors were applied to cells with final concentrations of 10 μ M. Dasatinib dissolved in DMSO was kindly provided by Henrik Daub.

Materials of inhibitor compounds: calyculin A was purchased from Millipore. Deltamethrin was purchased from Merck. U0126 was product from Promega. Na-orthovanadate, catalase and SB202190 were obtained from Sigma.

3.1.5. Cell stimulation, harvest and protein recovery

Cell stimulation in project 1:

In the double encoding SILAC experiment (Figure 4.1.2), Arg10 and Lys8 labeled Hepa1-6 cells were treated with 100 nM calyculin A, 0.5 mM pervanadate and 10 μ M deltamethrin for 10 minutes under 37 °C. Arg0 and Lys0 labeled cells were left untreated.

In the triple encoding SILAC experiment (Figure 4.1.17), cells were serum-starved for 24 hours. Arg0 and Lys0 labeled Hepa1-6 were left untreated. Arg6 and Lys4 labeled cells were

stimulated with 100 nM insulin for 5 minutes. Arg10 and Lys8 labeled cells were treated with 100 nM calyculin A and 0.5 mM pervanadate for 10 min (no deltamethrin).

Cell stimulation in project 2:

HeLa cells were serum-starved for 16 hours. Arg10 and Lys8 labeled HeLa cells were stimulated with 10 μ M kinase inhibitor (U0126 or SB202190) for 20 minutes. Then in the presence of the inhibitor compound, 150 ng/ml EGF was added to the medium for another 15 minutes. Arg6 and Lys4 labeled HeLa cells were stimulated with 150 ng/ml of EGF for 15 minutes. Arg0 and Lys0 labeled HeLa cells were left untreated.

To harvest cells, they were washed with ice cold PBS for two times and lysed in modified RIPA buffer containing 1% NP-40, 0.1% sodium deoxycholate, 150 mM NaCl, 1mM EDTA, 50 mM Tris, pH 7.5, 1 mM sodium orthovanadate, 5 mM NaF, 5 mM beta-glycerophosphate and protease inhibitors. 660 μ l of lysis buffer was added to each 15 cm dish. The lysates were kept cold and centrifuged at 17,000g for 15 minutes to pellet cellular debris.

To recover proteins by acetone precipitation (used in project 1 and 2): Supernatant was collected and mixed with four volumes of ice cold acetone to precipitate proteins. The cellular debris was digested with benzonase in urea buffer containing 6 M urea, 2 M thiourea and 10mM Hepes pH 7.5. Dissolved proteins were precipitated by adding four volumes of ice cold acetone. Precipitated proteins were collected by centrifugation and dissolved in urea buffer. Concentration of the dissolved proteins was measured by the Bradford method (A_{590nm}). Equal amounts of protein from each SILAC condition were mixed accordingly.

To recover proteins by methanol/chloroform precipitation (used in project 3): After centrifuge at 17,000g for 15 minutes (as described above), supernatant was collected and a Bradford method (A_{590nm}) was used to determine the protein concentrations. Equal amount of the proteins from the primary hepatocyte sample and Hepa1-6 sample were mixed, resulting in 100 μ g proteins in total. Protein mixtures were added with four volumes of methanol, one volume of chloroform and three volumes of distilled water in a sequential manner. The addition of each solvent was followed by a short vortex. After centrifugation of 20,000g for 1 minute, proteins were focused between organic and inorganic phases. The aqueous phase was discarded. Four starting volumes of methanol were added to the protein pellet followed by a short vortex. After spinning at 20,000g for 2 minutes, methanol was removed and the protein

pellet was air-dried. Precipitated proteins were redissolved in a buffer containing 6 M urea, 2 M thiourea, 10 mM Hepes, pH 7.5.

Materials: Insulin was purchased from Sigma Aldrich. EGF was from Millipore. Benzonase was from Merck. Protease inhibitors (complete tablets) were purchased from Roche Diagnostics. If not specified, chemicals were purchased from Sigma Aldrich. Solvent was obtained from Fluka, Merck, Riedel de Haen, and BioRad companies.

3.2 Proteome mapping

3.2.1 Sample preparation

Protein digestion

For protein in-solution digestion into peptides, the proteins were reduced with 1 mM dithiothreitol, alkylated with 5.5 mM iodoacetamide in dark, and digested for four hours with endoproteinase Lys-C (1/100 w/w). After diluting four times with Milli-Q water and adjusting to neutral pH with 20 mM ammonium bicarbonate, samples were digested overnight with sequencing grade modified trypsin (1/100 w/w). The digestion was quenched by adding trifluoroacetic acid (TFA) to reach pH <3. Chemicals and enzymes for protein digestion were dissolved in 20 mM ammonium bicarbonate. All incubation steps were performed at room temperature.

For protein in-gel digestion into peptides, protein samples were resolved on a NuPAGE 4%-12% Bis-Tris SDS-PAGE gel (Invitrogen). The gel was stained with the Colloidal Blue staining Kit (Invitrogen) for visualization. Gel slices were excised. Each gel slice was cut into 1 mm³ cubes, washed with 1:1 (v/v) 50 mM ammonium bicarbonate and 50% ethanol. 10 mM dithiothreitol was applied at 56°C for reduction and 55 mM iodoacetamide at 37°C for alkylation in dark. After wash, the gel pieces were dehydrated with 100% ethanol followed by rehydration with 12.5 ng/μl trypsin (Promega) in 50 mM ammonium bicarbonate. Trypsin digestion was performed overnight at 37°C. After digestion, supernatants were transferred to fresh tubes, and the remaining peptides were extracted by incubating gel pieces two times

with 30% MeCN in 3% trifluoroacetic acid (TFA), followed by dehydration with 100% MeCN.

Isoelectric focusing to separate peptide mixtures

Peptides were separated based on their isoelectric points in the Agilent 3100 OFFGEL Fractionator and the 3100 OFFGEL Low Res Kit, pH 3-10 according to the manufacturer. Peptides were focused for 20 kVh at maximum current of 50A and maximum power of 200 mW. Each peptide fraction was mixed 10 μ l solvent containing 30% MeCN, 5% acetic acid and 10% TFA. The resulting solution was loaded into C₁₈ reverse phase StageTips¹⁷⁶.

StageTip purification

StageTips^{176, 177} were prepared by punching out small discs of C₁₈ Empore filter using a 22 G flat-tipped syringe and ejecting the discs into P200 pipette tips. The C₁₈ Empore column was conditioned by methanol and equilibrated in 0.5% acetic acid, 0.1% TFA in water. Peptide samples were adjusted to pH<2.5 and forced through the C₁₈ Empore column. The column was washed once with 0.5% acetic acid, 0.1% TFA in water. Peptides were eluted from the StageTips by applying 80% MeCN, 0.5% acetic acid. Samples were dried in a SpeedVac to 3 μ l and mixed with equal volume of solvent containing 2% MeCN and 1% TFA. 5 μ l samples were applied for LC-MS/MS analysis.

Materials: Chemicals for the ‘in solution’ digestion were purchased from Sigma-Aldrich. Solvent was obtained from Fluka, Merck, Riedel de Haen, and BioRad companies. Endoproteinase Lys-C was obtained from Waco and sequencing grade modified trypsin was from Promega. Agilent 3100 OFFGEL Fractionator and 3100 OFFGEL Low Res Kit, pH 3-10 were purchased from Agilent. For producing StageTips, 3M High Performance Extraction Disks C₁₈ was obtained from Varian.

3.2.2 Mass spectrometric methods

C₁₈ RP-HPLC separation

The peptide mixture was separated by nanoscale C₁₈ reverse-phase liquid chromatography (Agilent 1200; Agilent Technologies, Waldbronn, Germany) coupled on-line to a 7-T LTQ-FT or LTQ-Orbitrap mass spectrometer (Thermo Electron, Bremen, Germany) as described¹⁷⁸. The HPLC is comprised of a solvent degasser, a nanoflow pump, and a thermostated micro-autosampler. The C₁₈ reverse-phase column, where the chromatographic separation of the peptides took place, is a 20-cm fused silica emitter with 75- μ m inner diameter (Proxeon Biosystems) packed in-house with methanol slurry of reverse-phase ReproSil-Pur C₁₈-AQ 3- μ m resin (Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) at a constant pressure (50 bar) of helium. Samples were picked up by the autosampler and injected to the C₁₈ reverse-phase column with a flow rate of 500 nl/min for the first 22 minutes. In the following, peptides were eluted at 250 nl/min with an actual separating gradient of 2-40% solvent (80% MeCN in water, 0.5% acetic acid) over 90 min. The eluate was on-line electrosprayed into the mass spectrometer via a nanoelectrospray ion source (Proxeon Biosystems, Odense, Denmark).

LC-MS/MS, Top 5 ion sampling

The mass spectrometers were operated in positive ion mode and employed a data-dependent automatic switch between MS and MS/MS acquisition modes. In the LTQ-FT setup, after accumulating a target value of 5,000,000 ions in the LTQ, a full scan was acquired in the FTICR analyzer with resolution $r=100,000$ at m/z 400. In LTQ-Orbitrap analysis, the target value was 1,000,000 for full scan in Orbitrap analysis at a resolution $r=60,000$ at m/z 400. In each cycle, full scan determined the five most intense ions from the range 300-1800 m/z . These five corresponding ions were next accumulated in the LTQ for sequential fragmentation. Total cycle time (full scan to full scan) was approximately 3 s. Fragmentation in the LTQ was induced by collision-induced dissociation with a target value of 5,000 ions. Former target ions selected for MS/MS were dynamically excluded for a period ranging from 60 to 300 s. The actual exclusion time was slightly adjusted according to samples.

The general mass spectrometric conditions were: spray voltage, 2.2 kV; no sheath and auxiliary gas flow; ion transfer tube temperature, 150-180°C; normalized collision energy

using wide-band activation mode; 35% for MS². Ion selection thresholds were: 500 counts for MS². An activation $q = 0.25$ and activation time of 30 ms was applied in MS² acquisitions.

Internal calibration using lock mass

To improve mass accuracy for identification, the “lock mass” option was utilized in both MS and MS/MS scans as described previously¹³⁰. This method makes use of the polydimethylcyclsiloxane (PCM) (Si(CH₃)₂O) ions generated in the electrospray process from ambient ion. The protonated PCM ions are with m/z of 445.120025, and in MS/MS mode with m/z of 429.088735. The lock mass was injected into the C-trap with a set “ion gain” achieving 10% of the target value of the full mass spectrum. The time of accumulation, isolation, and transfer into the C-trap of the lock mass was estimated to be a few ms and did not induce any obvious extension of ion cycle time.

Materials: Nanoscale C₁₈ reverse-phase liquid chromatography (Agilent 1100, Agilent 1200) was from Agilent Technologies (Waldbronn, Germany). Nanoelectrospray ion source was from Proxeon Biosystems (Odense, Denmark). Reverse-phase ReproSil-Pur C₁₈-AQ 3- μ m resins were from Dr. Maisch GmbH (Ammerbuch-Entringen, Germany).

3.3 Phosphoproteome mapping

3.3.1 Sample preparation

Immunoprecipitation of tyrosine phosphorylation proteins

Immunoprecipitation of tyrosine phosphorylated proteins using agarose-conjugated antibodies of 4G10 and pTyr-100 was performed as described before⁹³. Cells were lysed with modified RIPA buffer and centrifuged with 17,000g for 15 minutes, as described in section 3.1.5. The supernatant was used for immunoprecipitation. Briefly, non-specific binding proteins were pre-cleaned with Protein A (Sigma) beads and, following 2 hours incubation with 4G10 antibody, p-Tyr-100 antibody was added for an additional 4 hours. The incubation was performed at 4°C. Proteins were eluted from the beads with 3 times 1ml urea buffer. The

eluate was directly subjected to in-solution digestion. To collect the proteins not being eluted, the antibody beads were boiled in 70°C SDS-buffer for 10 minutes. SDS-PAGE was performed on a NuPAGE 4%-12% Bis-Tris gel (Invitrogen). In the triple encoding SILAC experiment in project 1 (Figure 4.1.17), the gel lane was cut into 4 large slices. Proteins were digested in-gel with trypsin, as described in section 3.2.1.

Crude fractionation of phosphopeptides by SCX

To separate phosphopeptides from non-phosphorylated peptides, the strong cation exchange (SCX) method was employed¹⁷⁹. 10-20 mg of proteins was digested in-solution and the peptide mixtures were adjusted by TFA to pH 2.7. Precipitates were cleared by centrifuging at 17000g for 10 min. Peptide solution was then loaded onto a 1 ml Resource S column (GE healthcare) connected to the Äkta Purifier chromatography system (Amersham Biosciences). The loading was performed in solvent A (5 mM KH₂PO₄, 30% MeCN, 0.1% TFA, pH 2.7) at a flow rate of 1 ml/min. During the loading step, flow-through solution was collected. Peptides bound to the column were separated with a linear gradient of 0-30% salt-containing solvent B (5 mM KH₂PO₄, 30% MeCN, 350 mM KCl, 0.1% TFA, pH 2.7). The separation lasted for 30 min. In total fifteen 2-ml fractions were collected by an automated fraction collector.

Phosphopeptide enrichment by TiO₂

Phosphopeptide enrichment by TiO₂ beads was essentially as described³¹ with slight modifications. TiO₂ beads were pre-incubated with 30 g/L 2,5-dihydroxybenzoic acid DHB in 80% MeCN and 0.1% TFA. From this 1:1 TiO₂ beads slurry, 10 µl was added to each sample and rotated end-over-end for 30 min. After one time wash with 1 ml 30% MeCN / 1% TFA and one time with 1 ml 50% MeCN / 1% TFA, bound peptides were eluted from beads with 200 µl NH₄OH in 40% MeCN (pH>10.5). Eluates were immediately neutralized in 30% MeCN / 3% TFA solvent. Samples were dried down almost to completion and reconstituted in 2% MeCN / 1% TFA for LC-MS/MS analysis.

Materials: 4G10 agarose conjugated antibody was purchased from Biozol. Phospho-tyrosine mouse mAb (P-Tyr-100) was purchased from New England. Protein A agarose conjugated beads were products of Sigma Aldrich. NuPAGE 4%-12% Bis-Tris gel and running buffer (MOPS) were obtained from Invitrogen. 1 ml Resource S column (GE healthcare) connected to the Äkta Purifier chromatography system (Amersham biosciences). Chemicals were products of Sigma Aldrich. Solvent was obtained from Fluka, Merck, Riedel de Haen, and BioRad companies.

3.3.2 Mass spectrometric methods

LC-MS/MS method, multi-stage activation

The mass spectrometers were operated in positive ion mode and employed a data-dependent automatic switch between MS and MS/MS acquisition modes. In the LTQ-FT setup, full scan mass spectra were acquired at a target value of 5, 000,000 ions with resolution $r=100,000$ at m/z 400. In the LTQ-Orbitrap setup, full scan was acquired at a target value of 1, 000,000 ions with resolution $r=60,000$ at m/z 400. For complex samples, a total cycle scan comprised of 3 mass ranges was applied: m/z 350-1050, 850-1850, and 350-1850. The top 5 most intense ions from the first 2 ranges were selected for fragmentation in the LTQ, whereas in the last range the top 7 most intense ions were selected. Fragmentation in the LTQ was induced by collision-induced dissociation with a target value of 5,000 ions.

For accurate mass measurement, the “lock mass” function was enabled for both MS and MS/MS scan modes. To improve the fragmentation of phosphopeptides, the multi-stage activation algorithm¹⁸⁰ in the Xcalibur software was enabled for each MS/MS spectrum. When a neutral loss of 97.97, 48.99, or 32.66 Thomson (Th) was detected following activation of the precursor ion, the neutral loss peptide fragments were further fragmented¹⁸⁰ to generate a variety of structurally informative fragments. Product ions produced during the initial activation of isolated precursor ion (MS/MS) and all subsequent neutral loss activations (pseudo MSⁿ) were simultaneously stored, resulting in “composite” mass spectra for peptide identification.

Former target ions selected for MS/MS were dynamically excluded for a period ranging from 60 to 300 s. The actual exclusion time was slightly adjusted according to samples. The general mass spectrometric conditions were essentially the same as described in section 3.2.2.

3.4 Mass spectrometric data analysis

MaxQuant is an in-house built mass spectrometric data processing software¹⁴⁰. It consists of three components: Feature detection and peptide quantitation (Quant.exe), Identification and validation (identify.exe), and Visualization (Viewer.exe). It develops an entire set of algorithms in detecting peaks, isotopes and SILAC partners, in quantifying relative peak intensities of SILAC partners, as well as in performing statistics to large scale proteomic datasets. Its strategy in using different charged states of the same peptide for non-linear re-calibration and those well identified peptides for global mass re-calibration has led to an 8-fold improvement in mass accuracy. This helps to rescue those identifications which would otherwise fall out of the required mass accuracy window and generally improves identification statistics. Different from conventional method, detection of SILAC partners in MaxQuant is performed before the identification and therefore they can be classified into groups such that heavy amino acid labeling can be treated as a fixed modification. This strategy significantly shortens the time for database search, where too many variable modifications often become a bottleneck in deriving results in time. Besides, the new strategy strictly constrains the possibility of mismatch.

3.4.1 Identification

Raw MS spectra were processed in Quant.exe and the derived peak list was searched with the Mascot search engine (Matrix Science, London, UK) against concatenated database combining forward database (the International Protein Index (IPI) protein database added with 27 commonly observed contaminants) and the reverse database (reversed sequences of all proteins from the forward database). For all mouse samples (Hepa1-6 cell line and primary hepatocytes) studied in this thesis, the IPI mouse protein database version 3.24 was employed, which records 52,326 proteins. For all human samples (HeLa and K562 cell line), IPI human protein database version 3.37 was used which contains 69,289 proteins.

Carbamidomethylation was set as fixed modification. As described above, MaxQuant classified SILAC partners according to labeling (i.e. light, medium, and heavy), and within each group the corresponding labeling was automatically set as fixed modification.

Variable modifications for proteome mapping dataset included oxidation (M), N-acetylation (protein), pyro (N-term QC). For phosphoproteome studies, variable modifications also include phospho(STY). Full tryptic specificity was required and up to three missed cleavages were allowed. Initial mass deviation of precursor ion and fragment ions were up to 10 ppm and 0.5 Da, respectively.

The derived identification lists of peptides and their assigned proteins were further processed in Identify.exe. Posterior error probability (PEP) was calculated to estimate the chance of random peptide match for each spectrum, given the Mascot score and peptide length. False discovery rate (FDR) denotes the percentage of false identification hits in the entire dataset. To achieve highly reliable identifications, the following standard operating procedures (SOP) were employed:

Proteome dataset: peptide PEP ≤ 0.1 , peptide FDR ≤ 0.01 , protein FDR ≤ 0.01 , peptide length ≥ 6 . To identify a protein, at least two peptides should be identified and one of them should be unique to this protein in the proteome.

Phosphoproteome dataset: peptide PEP ≤ 0.1 , peptide FDR ≤ 0.01 , protein FDR ≤ 1 , peptide length ≥ 6 . Phosphorylation analysis is performed at the peptide level.

3.4.2 Quantitation

For quantitation, peptide ratios were calculated according to the intensities of all 2D centroids from each of the SILAC forms. Linear line fitting to these intensities gave the slope as the desired ratio. During this calculation, element enrichment in ^{13}C , ^{15}N due to the SILAC labeling were taken into account. To represent the ratio of a peptide being quantified several times, the median value was chosen. To minimize the effect of outliers, protein ratios were calculated as the median of all SILAC pair ratios that belong to peptides contained in this protein. Intensity based significance value was also calculated to estimate the degree of biological regulation. Complementary to this fine-tuning method, a threshold fulfilling stringent requirement such as 2-fold change (Ratio ≥ 2 or Ratio ≤ 0.5) is also commonly used to derive biological regulation. The latter was applied in all the projects in this thesis.

3.4.3 Assign localization of the phosphate group

For each possible phosphorylation site, the localization post-translational modification (PTM) score was calculated by matching the observed b and y ions with the theoretical b and y ions³¹. Results are displayed in MaxQuant output tables. To accurately assign the phosphorylation sites, two processing steps were performed: entries with PTM scores that are lower than the maximum score minus five are ignored, and localization of PTM probabilities are required to be at least of 0.75.

3.4.4 Phosphoproteomic dataset stored in PHOSIDA database

All the phosphoproteomic datasets from project 1 and 2 are stored in the PHOSIDA database as a public resource. The data will become available along with the publications. PHOSIDA routinely performs motif check (see section 3.5.2), structural analysis (see section 3.5.3), phosphorylation site prediction (see section 3.5.4), and evolutionary comparison¹⁸¹.

3.5 Bioinformatic analysis

3.5.1 Gene Ontology and KEGG enrichment analysis

Routine GO analysis

Cytoscape along with its Plug-in Bingo 2.0¹⁸² was used to analyze the distribution of experimental datasets among various protein groups, and to identify significantly overrepresented biological functions of the proteins.

The Gene Ontology (GO) annotations of proteins were compared with the ones of a reference proteome (e.g. identified proteins vs. the entire protein database, or a subset of the identified proteins vs. the overall identified proteins). To assign corresponding GO identifiers to each IPI entry, the Gene Ontology Annotation database was used. The hypergeometric test and the Benjamini & Hochberg False Discovery Rate correction were performed to derive overrepresented functions¹⁸². A probability value of 0.05 was considered significant.

Enrichment analysis based hierarchical clustering

In project 3, *Proteomic phenotyping to assess differences between transformed and non-transformed mouse liver cells*, the quantified proteome was divided into 5 quantiles corresponding to probability cutoffs of 0, 0.15, 0.25, 0.75, 0.85, and 1. The enrichment analysis for GO biological process and cellular component were done separately for these quantiles with respect to the whole quantified proteome by conditional hypergeometric test available in the GOstats package¹⁸³ in the R statistical environment¹⁸⁴. For hierarchical clustering we first collated all the categories obtained after enrichment along with their p-values, and then filtered for those categories which were at least enriched in one of the quantiles with p-value < 0.05. This filtered p-value matrix was transformed by function $x = -\log_{10}(p\text{-value})$. Finally these x values were transformed to z -score for each GO category by using the transformation:

$$\frac{x - \text{mean}(x)}{\text{sd}(x)}$$

These z -scores were then clustered by one-way hierarchical clustering using “Euclidean distance” as distance function and “Average Linkage clustering” method available in Genesis¹⁸⁵. KEGG pathway enrichment analysis was done in the same way, except that the hypergeometric test was employed.

3.5.2 Phosphorylation - motif check

The known human kinase motifs were used to access the mouse phosphorylation dataset to identify possible matching kinase substrates¹⁸¹. χ^2 -Test was employed to perform the statistical evaluation. The applied χ^2 -Test checks whether the observed number of phosphosites that match with a given kinase motifs exceeds the number of expected sites to a statistically significant extent. To derive the number of expected motif matching phosphosites, the chance was estimated for each kinase motif to match with a given phosphosite according to the amino acid composition of the motif and the relative frequencies of each amino acid in the entire mouse proteome. This approach was applied to those phosphorylation sites that could be clearly assigned within the phosphorylated peptide sequence (class I sites).

Furthermore, Motif-X¹⁸⁶ was used to derive potentially new motifs in-silico and to confirm the observations of the χ^2 -Test. A probability value of 0.0001 was considered significant. In

addition, a minimum occurrence of 20 was required to derive a significant consensus sequence. This approach was performed residue-specific on all class I sites using the entire mouse proteome as a background model.

3.5.3 Phosphorylation - structural analysis

In order to check the structural constraints of phosphorylated residues in the mouse liver cells used in this thesis, the secondary structure and solvent accessibility prediction tool SABLE 2.0¹⁸¹ was used. The predicted structural conditions of each phosphosite were stored in PHOSIDA.

3.5.4 Phosphorylation - phosphorylation site predictor

A support vector machine (SVM) was trained separately on unambiguously identified phosphorylation sites¹⁸¹. The essential feature of each phosphorylation site that was used as input for this machine learning approach was the raw sequence: the algorithm was mainly trained on the phosphorylated residue along with its surrounding sequence (+/- 6 residues). To generate a negative set of the same size, sites from mouse proteins that were not, to date, detected to be phosphorylated were randomly chosen. The positive and the negative datasets were split into a training set (90%) and a test set (10%). Parameters C and σ were then optimized by varying them from 2^{-10} to 2^{10} in multiplicative steps of two on the basis of a five-fold cross validation on the training set. The optimal model for each set of each phosphorylated amino acid was obtained separately using the radial basis function (RBF).

4 Results and Discussions

4.1 Project 1 - Quantitative phosphoproteome analysis of a liver cell line reveals specificity of phosphatase inhibitors

This work is included in a manuscript that is accepted for publication:

Quantitative phosphoproteome analysis of a liver cell line reveals specificity of phosphatase inhibitors

Cuiping Pan, Florian Gnad, Jesper V. Olsen and Matthias Mann

Phosphorylation is chemically stable under physiological conditions. However, enzyme-catalyzed phosphorylation and dephosphorylation can occur at ice-temperatures and in disrupted cells, e.g. cell lysates. Chemical compounds such as protease inhibitors and phosphatase inhibitors are added during sample preparation, to prevent loss of activity and dephosphorylation of the proteins *ex vivo*. Cell-permeable phosphatase inhibitors can also be used on living cells to non-specifically increase levels of phosphorylation, for example in phosphoproteomics. Some of these inhibitor compounds are known to target a broad spectrum of phosphatases. Here we wanted to use these broadband phosphatase inhibitors to measure and quantify a mouse liver cell line (Hepa1-6) phosphoproteome in greatest depth. Additionally, the strategy of metabolically labeling the Hepa1-6 using SILAC enabled us to compare the specificity of these phosphatase inhibitor compounds. This can be achieved by comparing the number of regulated tyrosine phosphorylation sites and the number of regulated Ser/Thr phosphorylation sites against the overall identified phosphoproteome.

Furthermore, by introducing a biological stimulus into an additional SILAC state we were able to determine an upper limit to the stoichiometry of phosphorylation sites.

The following three inhibitor compounds were used in this study. Sodium-pervanadate (Figure 4.1.1 A) is a broadband inhibitor for protein tyrosine phosphatases (PTPs). Calyculin A (Figure 4.1.1 B) targets PP1 and PP2A, the two largest groups of the serine and threonine phosphatase superfamily. Deltamethrin targets PP2B phosphatases.

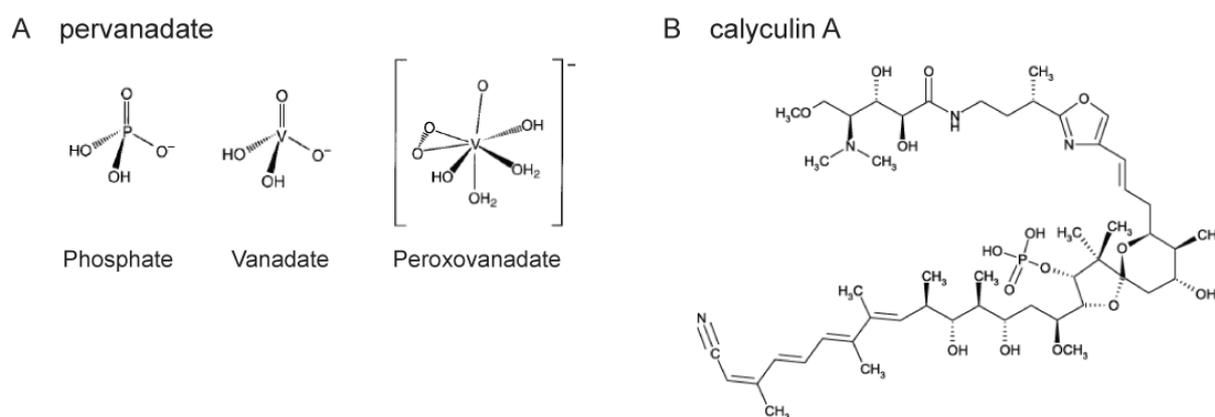


Figure 4.1.1 Chemical structures of phosphatase inhibitors. Panel A displays the monoperoxo form of pervanadate in comparison to the phosphate group and vanadate. Panel B displays the chemical structure of calyculin A.

Pervanadate is a general term for various reaction products of hydrogen peroxide and orthovanadate. These peroxovanadium compounds are general PTP inhibitors. Though analog to phosphate group, their mechanism of inhibition is not simply as competitor with phosphate, but rather it is suggested that they oxidize the catalytic cysteine residue in the PTPs¹⁸⁷. Calyculin A is a natural product isolated from marine sponge. It possesses a spiro ketal skeleton. Although its exact mechanism of action is not clear, it has been demonstrated to be highly potent in inhibiting pSer/Thr phosphatases^{188, 189}. Deltamethrin, a pyrethroid ester insecticide, has proven to be a specific inhibitor for PP2B¹⁹⁰. All these three selected inhibitors are membrane permeable and therefore suitable for intact cell treatment.

We employed SILAC-based phosphoproteomics strategy in this study (Figure 4.1.2). One cell population was labeled with normal ('light') arginine and lysine and another cell population with 'heavy' $^{13}\text{C}_6^{15}\text{N}_4$ -arginine and $^{13}\text{C}_6^{15}\text{N}_2$ -lysine (Arg10 and Lys8). After complete

incorporation, the heavy labeled cells were treated for 10 minutes with a mix of the three commonly used phosphatase inhibitors: 100 nM calyculin A, 0.5 mM pervanadate and 10 μ M deltamethrin, in order to block phosphatases as broadly as possible. We lysed treated and control cells, precipitated proteins, mixed them in a one to one ratio and enzymatically digested the resulting 20 mg protein mixtures. Peptides were separated by strong cation exchange (SCX) and the resulting fractions enriched for phosphopeptides using TiO_2 beads in the presence of DHB (section 3.3.1). Phosphopeptides were analyzed by high resolution mass spectrometry on a linear ion trap Fourier Transform instrument (LTQ-FT), using multistage activation. In all, 16 LC MS/MS runs were performed, using conditions described in section 3.3.2. Data was analyzed as described in section 3.4, requiring a False Discovery Rate (FDR) for phosphopeptide identification of less than 1%.

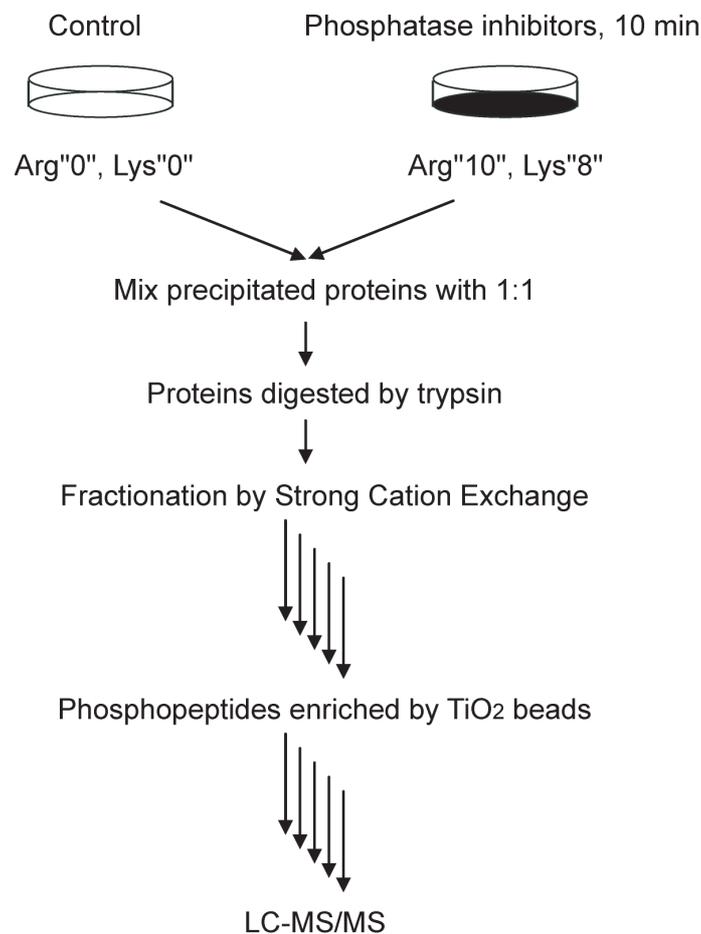


Figure 4.1.2 The strategy of SILAC-based quantitative phosphoproteomics to elucidate the efficiency of broadband phosphatase inhibitors in retaining phosphorylation.

4.1.1 Evaluation of the phosphoproteomics technique

Crude fractionation of phosphopeptides by SCX

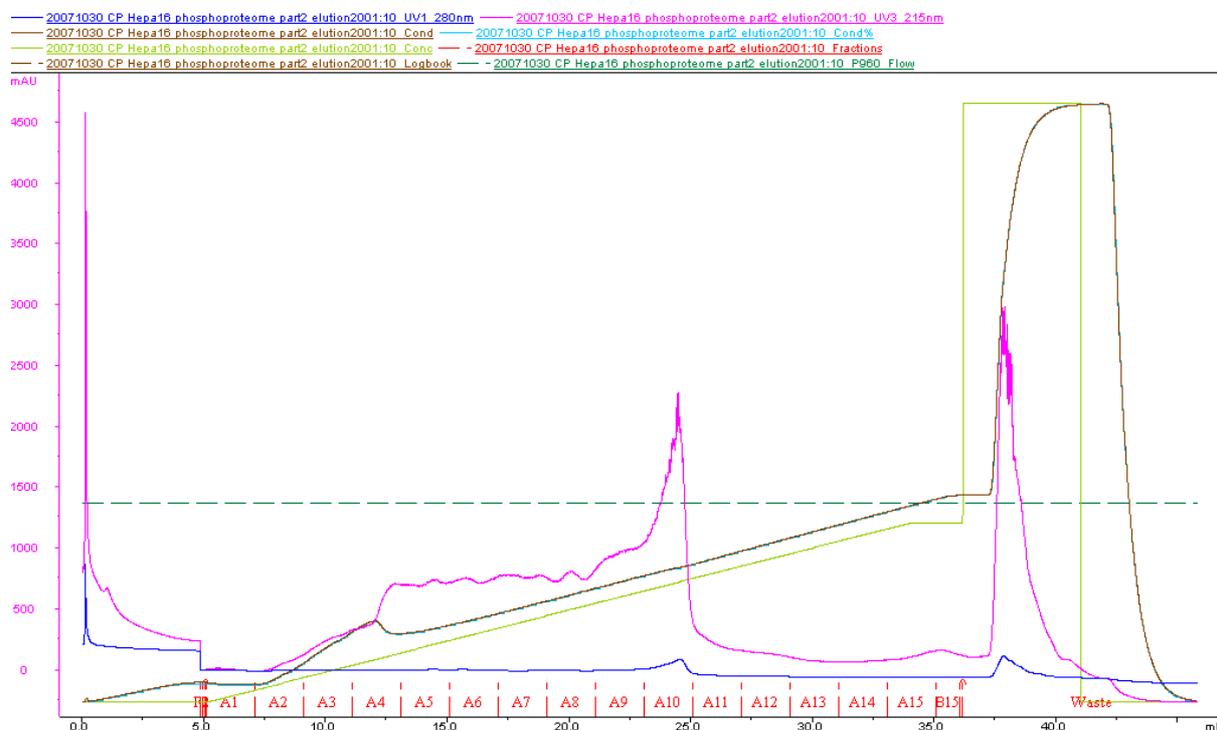


Figure 4.1.3 Phosphopeptide elution profile of SCX. X-axis represents time and y-axis represents UV absorbance at 225 nm (magenta). The green curve indicates salt concentration. It directly correlates with the conductivity curve marked in brown. The blue curve displays UV absorbance at 280 nm. Other parameters are written at the top. Note the fraction numbers are adjacent to the x-axis, namely A1 to B15..

A total of 20 mg of tryptic peptides were divided into four equal units and fractionated by SCX separately. The resulting fractions were combined correspondingly. In each round of sample loading, liquids which did not bind to the column (“flow through” fraction) were collected. The peptides that bound to the column were eluted by linearly increasing salt concentration. All four elution profiles were very similar. One of them is shown in Figure 4.1.3. Generally the elution profile is not sharp, suggesting that only a crude fractionation is achieved. UV absorbance indicates significantly fewer peptides at the late phase of elution (fraction A11-B15).

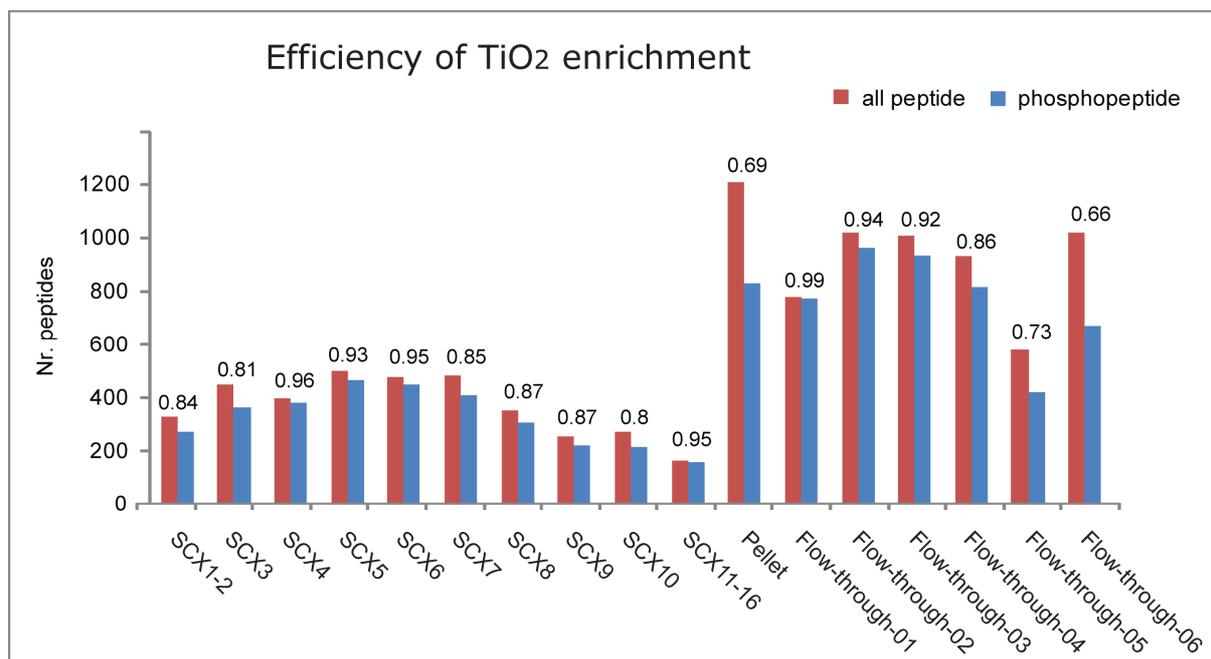
The efficiency of phosphopeptide enrichment by TiO₂ spheres

Figure 4.1.4 Phosphopeptide enrichment efficiency of TiO₂ sphere. SCX fractions are marked in x-axis. Y-axis presents the number of peptide identification. On top of each column pair marks the enrichment efficiency, which was obtained by dividing the number of phosphopeptides by all peptides being identified in that fraction(s). Calculation is based on evidence.txt output from MaxQuant.

The resulting fractions from SCX (Figure 4.1.3) were incubated with TiO₂ sphere for phosphopeptide enrichment. The early fractions and the late fractions were combined before incubation. The flow-through fraction, which contained mainly multiple phosphorylated peptides, as well as the pellet sample from the cell harvesting step (section 3.1.5) were also incubated with TiO₂. Because we estimated that the flow-through fraction might contain a rich amount of phosphopeptides, six times of TiO₂ enrichment were performed in sequence.

The enrichment worked well, with an overall efficiency of 0.85, defined as the fraction of phosphopeptides to all peptides: 100% enrichment efficiency gives the number of 1. The enrichment efficiency is displayed in Figure 4.1.4. Out of 17 samples, six samples achieved an efficiency equal to or above 0.95. It is noteworthy that the flow-through and pellet samples contained many more phosphopeptides than the SCX fractions, indicating that a large portion of the peptides did not bind to the SCX column. In fact, the six flow-through samples contributed 37% of overall identified non-redundant phosphopeptides (1,996 out of 5,457).

Further investigation revealed a significant overlap between adjacent flow-through samples. Interestingly, the first two of them constituted 1,222 phosphopeptides, 61% of those identified in all the combined six flow-through samples. The number of additional identified phosphopeptides levels off with sequential TiO₂ enrichment (Figure 4.1.5). This could be caused by abundant phosphopeptides saturating the beads or less likely that we readily exhaustively recovered all phosphopeptides from these hepatocytes. Therefore further fractionation on the SCX flow-through sample should effectively enlarge the identification numbers. A promising method to do this would be strong anion exchange but this was not pursued here to keep the number of fractions to be analyzed manageable.

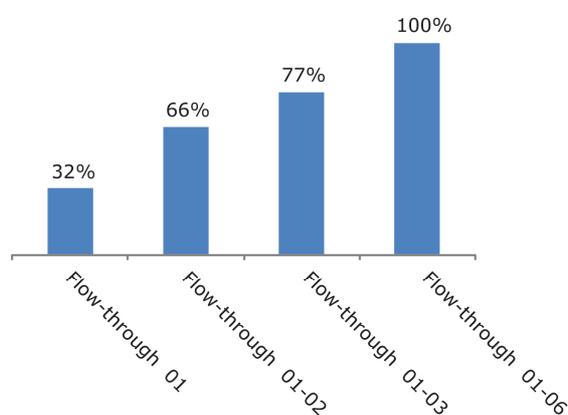


Figure 4.1.5 Phosphopeptide identifications in flow-through samples. Calculation is based on non-redundant phosphopeptides.

HPLC chromatogram

Depicted in Figure 4.1.6 are two HPLC chromatograms from one SCX fraction and one flow-through fraction. DHB introduced chemical noise into the mass spectra. The disturbance is obvious during sample loading, i.e. the first 22 minutes. However, when phosphopeptides started to elute their signal ‘suppressed’ the chemical noise.

Phosphopeptides are typically of low abundance. In our mass spectrometric data, SCX fractions derived from 20 mg starting material typically required around 150 ms fill time in each survey scan to fill the trap with one million ions. Due to higher phosphopeptide abundance, the typical ion fill time of the flow through fraction from SCX loading was only 50 ms.

Results and Discussions

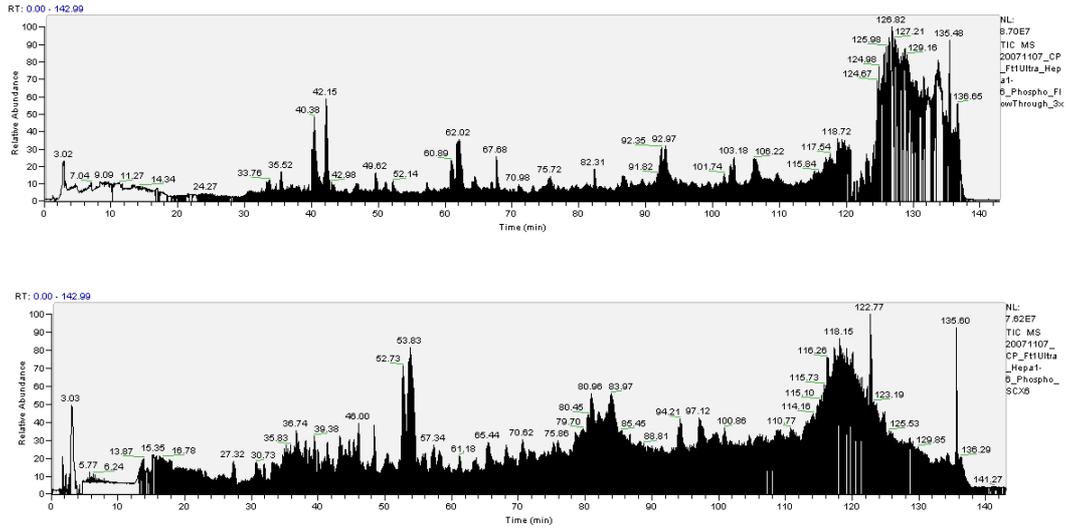


Figure 4.1.6 HPLC elution profiles of a complex phosphopeptide mixture.

Mass spectra: accurately pinpointing the localization of the phosphate group

Phosphopeptide identification was performed with the software MaxQuant. Three MS/MS spectra are shown as examples in Figure 4.1.7 to Figure 4.1.9.

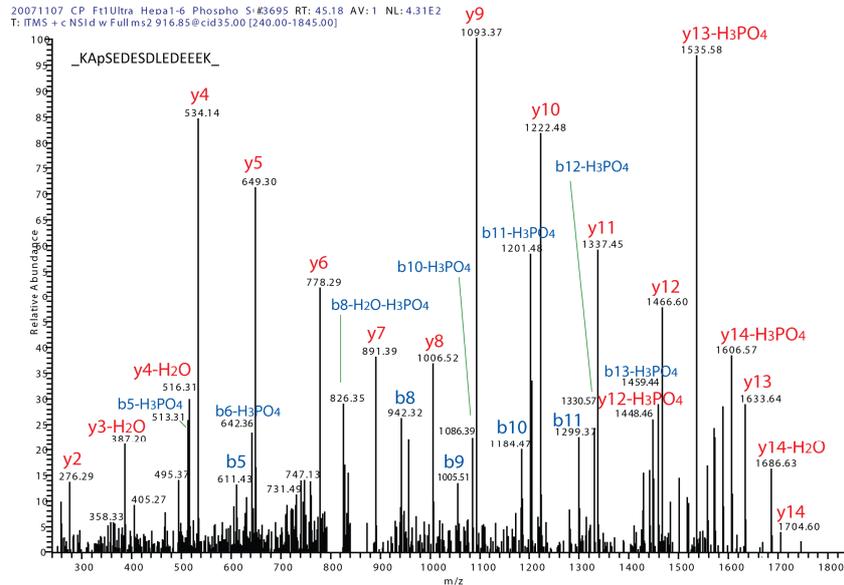


Figure 4.1.7 MS/MS spectrum to identify the pSer 668 in DNA replication licensing factor MCM3. The $b_5\text{-H}_3\text{PO}_4$, $b_6\text{-H}_3\text{PO}_4$, $y_{13}\text{-H}_3\text{PO}_4$ and $y_{14}\text{-H}_3\text{PO}_4$ fragments clearly pinpoint the location of phosphate group to S668.

Results and Discussions

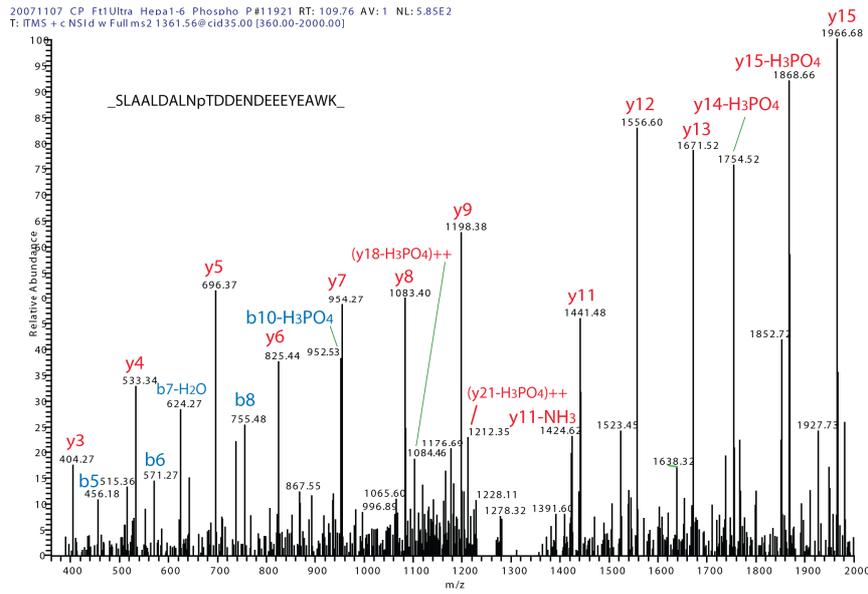


Figure 4.1.8 MS/MS spectrum to identify the pThr 267 in microfibrillar-associated protein 1. The b_{10} -H₃PO₄, y_{14} -H₃PO₄ and y_{15} -H₃PO₄ fragments clearly pinpoint the location of phosphate group in T267 but not another serine or tyrosine residue.

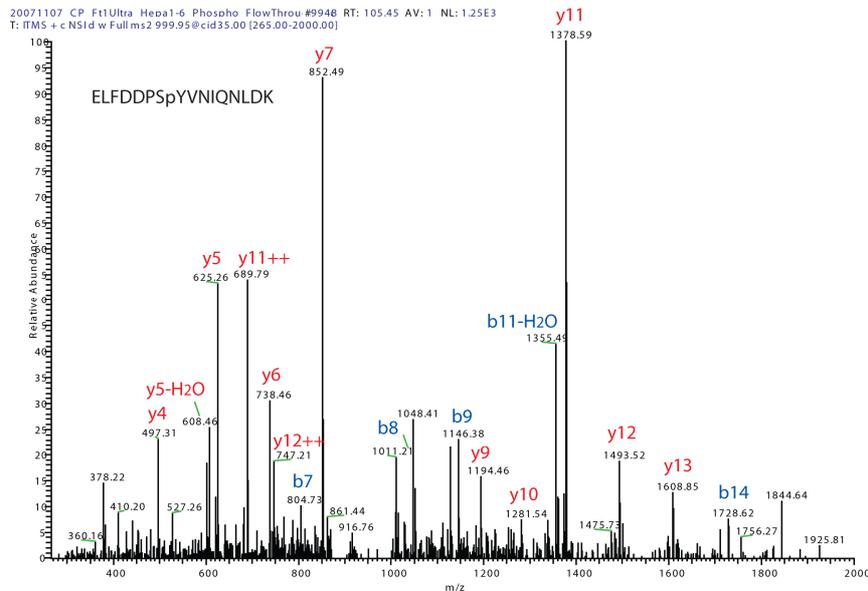


Figure 4.1.9 MS/MS spectrum to identify the pTyr 427 in SHC-transforming protein 1. The abundant y_7 and y_9 ions strongly indicate that the phosphate group locates to the tyrosine residue but not the serine residue.

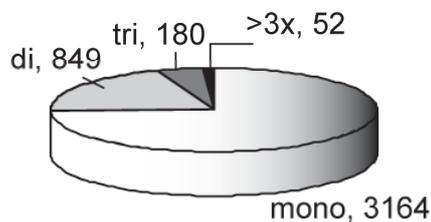
4.1.2 Phosphoproteome of the mouse liver cell line Hepa1-6

4.1.2.1 Overall identified phosphorylation sites and their distributions

The combination of phosphatase inhibitor compounds in this experiment was close to the maximum dose and duration tolerated before cells started to detach and show signs of apoptosis. Normally growing cells do not significantly phosphorylate many of the important substrates that are phosphorylated in specific situations, such as after a given stimulus or at a specific point in the cell cycle. Phosphatase inhibitor treatment may enhance these phosphorylation sites and thereby make them analyzable by MS. In this way we could characterize the phosphoproteome in greater depth. Also, this quantitative strategy allows us to determine if the peptide is present – at least in a small amount – in normal cells, i.e. without phosphatase inhibitor treatment.

A total of 1,808 phosphoproteins, 3,430 phosphopeptides, and 4,253 Class I phosphorylation sites were sequenced and identified. The majority of phosphopeptides were singly phosphorylated (75%), but a substantial fraction were either doubly (20%), triply (4%), or more highly (1%) phosphorylated (Figure 4.1.10 A). The distribution between phosphotyrosine (pTyr), phosphothreonine (pThr) and phosphoserine (pSer) was close to that observed in another large-scale phosphoproteomics study of human epithelial cells (HeLa) (Figure 4.1.10 B)³¹.

A



B

| Sites | Class I | Percentage |
|-------|---------|------------|
| All | 4253 | 100 |
| pS | 3733 | 87.8 |
| pT | 437 | 10.3 |
| pY | 83 | 1.9 |

Figure 4.1.10 (A) distribution of mono- and multiple phosphorylated peptides. (B) Frequency of class I sites, which are sites with the highest localization probability.

4.1.2.2 Characterization of the phosphoproteome of a mouse live cell line

The phosphoproteome is distributed among diverse protein classes, functions and subcellular locations.

Gene Ontology (GO) analysis assigned the phosphoproteome to cellular components, biological process and molecular functions (Figure 4.1.11). Given the fact that phosphorylation is involved in various cellular events, it is not surprising that these phosphoproteins are distributed among many different categories and functions. There are a large number of regulatory proteins in the phosphoproteome – for example 51 phosphorylated transcription factors and 121 phosphorylated protein kinases and 28 phosphatases.

The phosphoproteome is overrepresented in protein binding and underrepresented in mitochondrial and secreted functions.

In GO, we next analyzed which protein functions are overrepresented in the detected phosphoproteome compared to the entire mouse protein database. The results indicated that the most significantly overrepresented biological functions of phosphorylated proteins are associated with binding to targets ranging from transcription factors to ATP. Kinase binding activity, for example, is significantly overrepresented ($p= 8*10^{-7}$). Functions that are related to general kinase activities, translational activation, and transcriptional regulation also proved to be significant. Conversely, mitochondria and secreted proteins were significantly underrepresented in the phosphoproteome similar to what we had already seen in the HeLa phosphoproteome³¹.

Results and Discussions

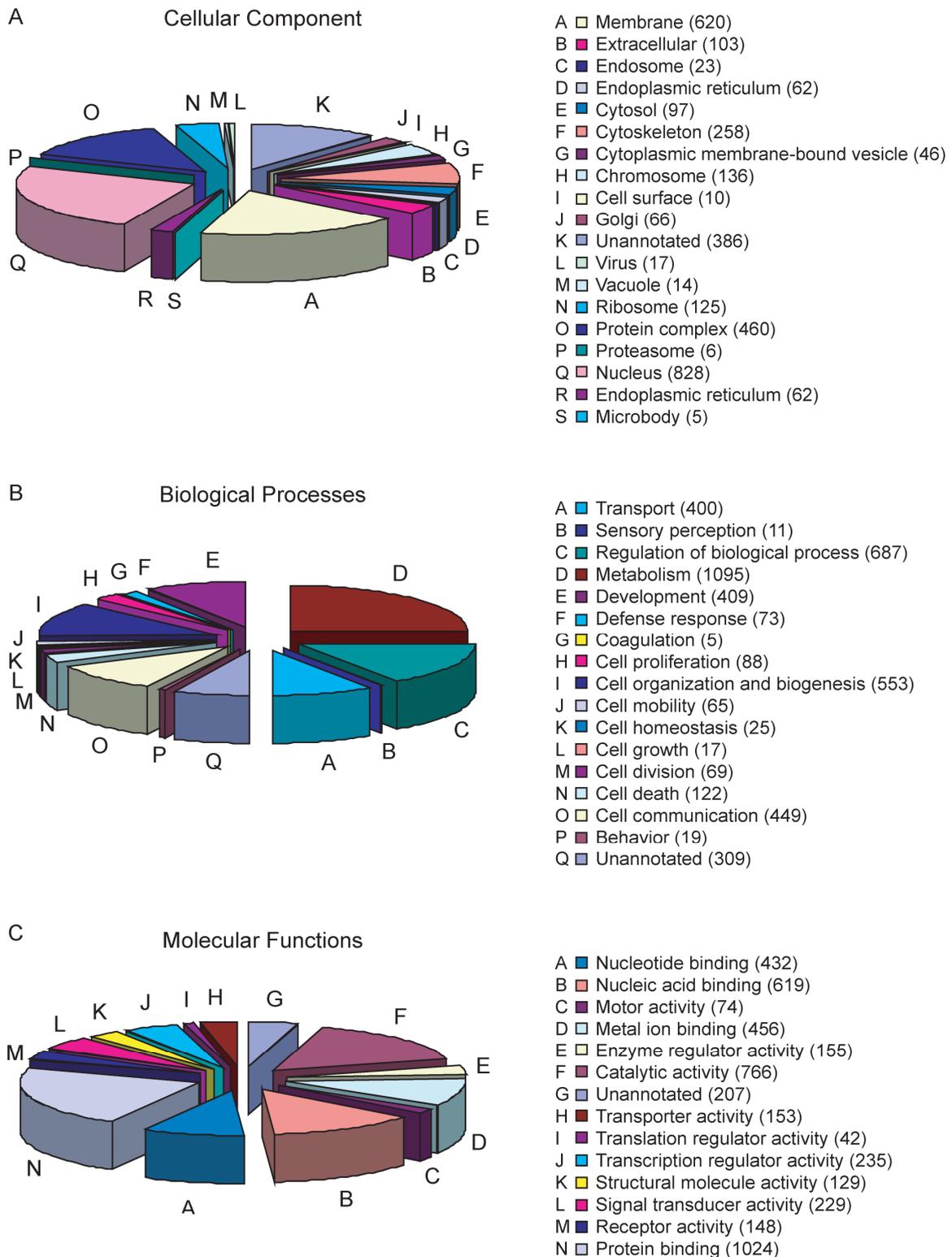


Figure 4.1.11 Overview of the proteins categories found in the phosphoproteome. Some regulatory protein classes such as protein binding, kinase activities, as well as transcriptional and translational regulations are strongly represented. Analysis was performed using Gene Ontology (GO).

Phosphorylation largely localized to coil regions of protein surfaces.

The solvent accessibility and secondary structure prediction by the bioinformatic software SABLE 2.0 revealed that the predominant localization of mouse phosphosites is in coil regions on the protein surfaces (Figure 4.1.12). This result is in concordance with the previous observation on structural constraints of phosphosites in the human proteome¹⁸¹.

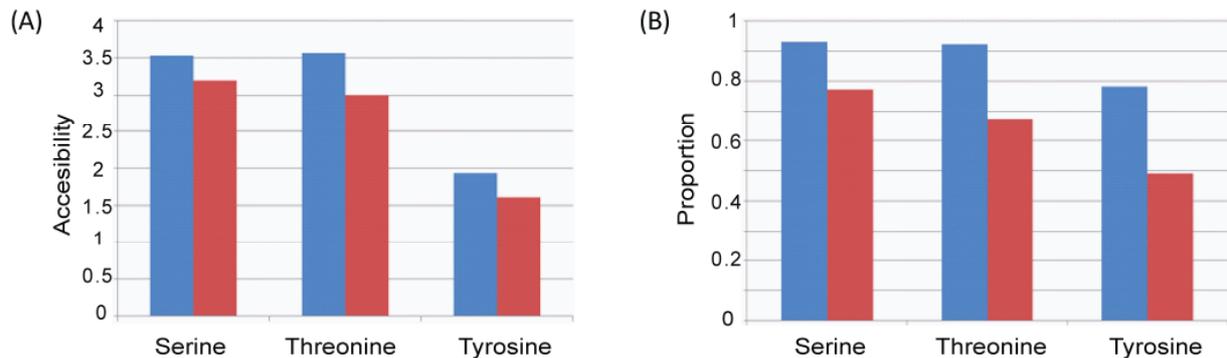


Figure 4.1.12 (A) Average solvent accessibility of phosphorylated residues is significantly higher than that of non-phosphorylated residues. The relative accessibility prediction assigns a value between 0 (fully buried) and 9 (fully exposed) to each residue. (B) Phosphorylated residues are more frequently located in loops and coils than the non-phosphorylated residues. Calculations were performed in SABLE 2.0. Blue: phosphorylated residue; red: non-phosphorylated residue.

Overlap between the mouse live cell line phosphoproteome and SwissProt phosphorylated proteins.

In this experiment, 1,033 phosphorylated proteins were identified that contain class I phosphorylation sites and have a corresponding SwissProt accession numbers. The current SwissProt release (54.7) reports 3401 proteins that are phosphorylated in mouse, disregarding method of experimental detection and including computational prediction by similarity. In total, 864 proteins overlapped between these two datasets (Figure 4.1.13). This is a striking overlap, given that there are more than 50,000 protein entries in the mouse IPI database. This observation implies that the core phosphoproteome, at least at the level probed here, is quite well conserved between different mammalian cell types and tissues. However, at the

phosphosite level, our dataset has substantial novelty as more than half (1,428) of 2,590 class I sites, whose assigned proteins show SwissProt accession numbers, is novel compared to what is currently recorded in SwissProt. In total, SwissProt contains 11,304 residues that are phosphorylated on mouse proteins.

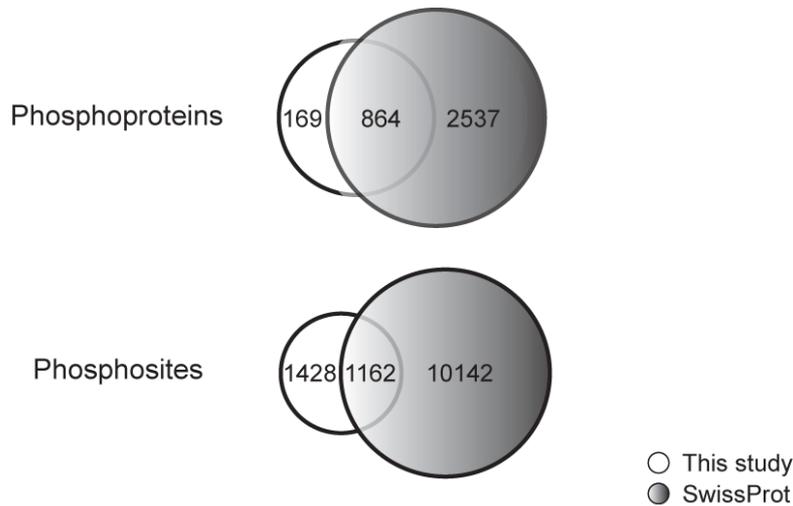


Figure 4.1.13 Overlap of phosphorylation sites found in this study with phosphorylation sites in the SwissProt database.

Overlap between the mouse live cell line phosphoproteome and mouse liver tissue phosphoproteome.

We were also interested in determining the overlap in phosphorylation sites between our liver cell line and liver tissue. Gygi and coworkers recently published a large-scale study of the mouse liver phosphoproteome which contains over 5,600 phosphorylation sites¹⁹¹. We extracted those phosphorylation sites and compared them to our dataset. Interestingly, this overlap was much smaller than the overlap with the phosphorylation sites in SwissProt. About half of our phosphoproteins were also detected in the tissue study, and about 30% of the phosphorylation sites on those proteins were identical. At this point, it is difficult to assess whether this discrepancy is due to differences between cell lines and tissues, or due to technical factors – both studies identified roughly equally large proteomes but neither is exhaustive.

Consensus sequences (motif) of kinase substrates.

Each of the phosphorylation sites identified in our dataset is the substrate of one or more kinases. Through motif analysis we matched our sites to the known substrate specificities of 33 human kinases. We used the published human kinase motifs (www.phosida.com) because the mouse ones are not known and kinase substrates are generally assumed to be well conserved between human and mouse. For each kinase motif, we determined the number of phosphorylation sites that matched the pattern, the number of sites expected to match this pattern by chance and the statistical significance of any overrepresentation using the χ^2 -test. The main finding of the consensus sequence analysis is that phosphorylation sites of the mouse proteome match significantly with specified human kinase motifs with only a few exceptions such as the motif of the NEK6 kinase (Appendix 2). The number of phosphosites that match with given motifs of the protein kinase A (PKA) is ten times higher than one would expect by chance, for example. These results were confirmed by the Motif-X algorithm¹⁸⁶ (Appendix 3).

4.1.2.3 Assessment of our phosphoproteomics technology*More than three orders of magnitude in phosphopeptide intensities.*

To our knowledge the dynamic range of phosphopeptide detection has not been described before. We determined the apex of the intensity over the LC MS peak for the isotope cluster of each phosphopeptide. Figure 4.1.14 shows that the identified phosphopeptides follow a Gaussian intensity distribution on a logarithmic x-axis, which spans between three to four orders of magnitude. This is the same dynamic range than that of unmodified peptide detection on our LTQ-FT instrument¹³¹, and for the LTQ-Orbitrap¹⁹² and probably reflects technical rather than biological properties.

Also indicated in the figure is the distribution after phosphatase inhibitor treatment. This curve is shifted by about a factor two relative to the untreated population, indicating that identification of lower abundance phosphopeptides from untreated cells was indeed enabled by this phosphatase strategy. This trend is especially striking at the left side of the graph, where the abundance bins contain very few 'heavy labeled' peptides. These 'light'

phosphopeptides were only identified because of their more abundant, SILAC-counterparts that had been phosphatase inhibitor stimulated.

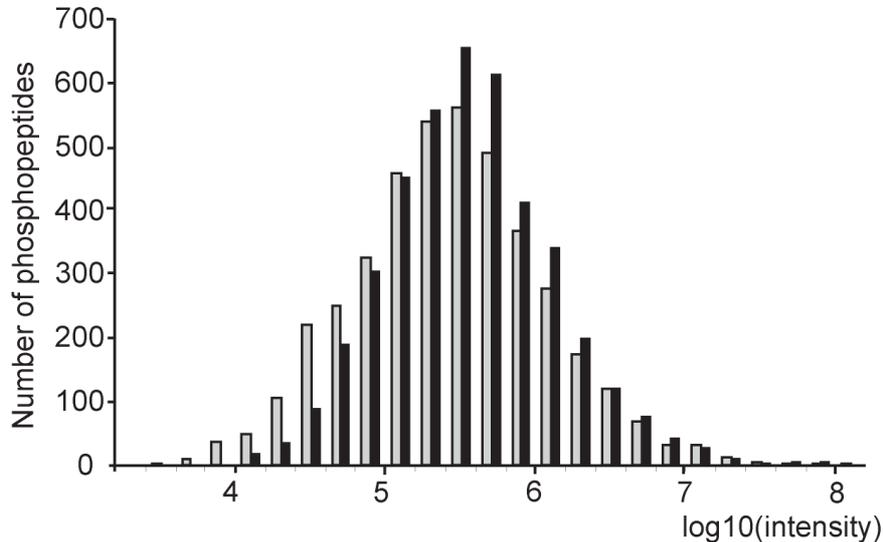


Figure 4.1.14 Number of phosphopeptides in different intensity bins. Just like the proteome, the phosphoproteome follows a roughly Gaussian distribution on a logarithmic scale. The light shaded bars represent the phosphopeptides from the light labeled SILAC cells (untreated) and the black bars the phosphopeptides from the heavy labeled SILAC cells (phosphatase inhibitor treated).

Phosphorylation correlates with protein abundance.

We next compared the abundance of the phosphoproteome of 1,808 proteins to the abundance of the proteome of the same cell type (data from Project 3). The proteome (4,089 proteins) was binned into abundance classes by means of the added peptide intensities (Figure 4.1.15 A). We then binned the phosphoproteins into the same abundance classes (Figure 4.1.15 B). Comparison of panels A and B reveals that the phosphoproteome roughly follows the abundance curve of the proteome, suggesting that to some degree phosphorylation occurs evenly among proteins without bias with respect to abundance. Thus, hyper-phosphorylation on low abundant proteins, or hypo-phosphorylation on high abundant proteins seemingly is not the general phenomena in cells. This comparison also suggests that we are not substantially biased against low abundance proteins – at least not compared to a proteome analysis.

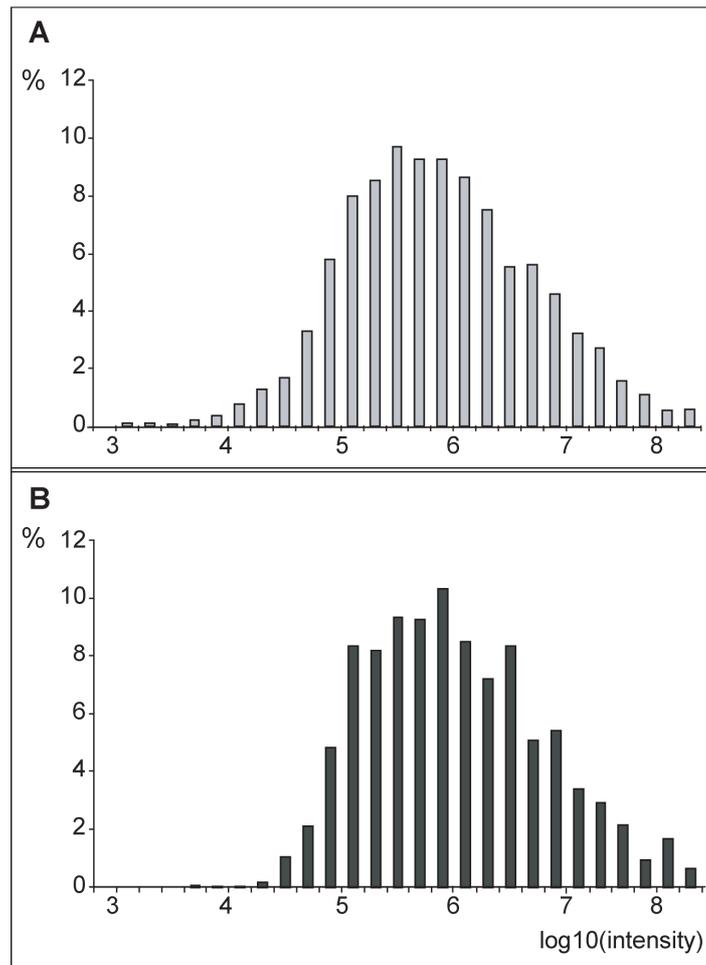


Figure 4.1.15 Proteome and phosphoproteome abundance. (A) Abundance of 4,089 proteins in the Hepa1-6 proteome (data from project 3). (B) Percentage of proteins in (A) for which at least one phosphorylation site was detected. Note that the left hand tail in this curve is depressed even though phosphoproteome measurement enriches for low abundance proteins. This is caused by the fact that the chance to detect the corresponding protein in a proteome measurement for a low abundance phosphopeptide is small.

4.1.3 Effects of calyculin A, deltamethrin, and pervanadate

4.1.3.1 Regulation of phosphopeptides by the inhibitor compound mixtures

We next asked how many phosphopeptides showed a fold-change due to the phosphatase inhibitor treatment (Table 4.1.1). Surprisingly, only 27% of the peptides were induced more than two-fold by this treatment. Note that some phosphorylation sites (8%) decreased instead of increased after phosphatase inhibitor treatment. This is due to the fact that phosphorylation

can both activate and deactivate target proteins and its downstream signaling cascade. For example, dephosphorylation in glycogen synthase is required for its activation.

| Fold change | Nr. | % |
|----------------------|------|------|
| ≤ 0.5 | 570 | 8.3 |
| 0.5 - 2.0 | 3853 | 56.3 |
| ≥ 2 | 1203 | 17.6 |
| ≥ 4 | 797 | 11.7 |
| ≥ 8 | 416 | 6.1 |
| All phosphopeptides* | 6839 | 100 |

Table 4.1.1 Overall regulation by phosphatase inhibitors. The star (*) indicates that the total number of phosphopeptides is calculated by separately summing up all detected phospho-states for the same peptide.

4.1.3.2 Overall regulation of the inhibitor compound mixtures

As noted, not all phosphosites were equally affected by phosphatase inhibitors. We plotted the fold-change of phosphorylation sites for each amino acid separately in Figure 4.1.16. As tyrosine phosphorylation is much less abundant compared to serine and threonine phosphorylations, we performed immunoprecipitation (IP) to enrich pTyr containing peptides using 4G10, a pTyr specific antibody, and obtain 480 pTyr sites (section 4.1.4). The data displayed in Figure 4.1.16 A come from this 4G10 IP experiment.

It is immediately apparent from the figure that tyrosine phosphorylation sites are dramatically affected by phosphatase inhibitor, while this is much less the case for serine/threonine. For pTyr, the majority of sites are up-regulated at least two-fold (70%). This is presumably due to the broad effect of pervanadate on all PTPs combined with the low level of tyrosine phosphorylation in untreated resting cells. For phosphothreonine only 41% of the sites were up-regulated at least two-fold and for phosphoserine the number is only 26%. This is a surprisingly low number considering that the inhibitors we used are generally employed in cell biology, and are usually thought to block most phosphatase activity.

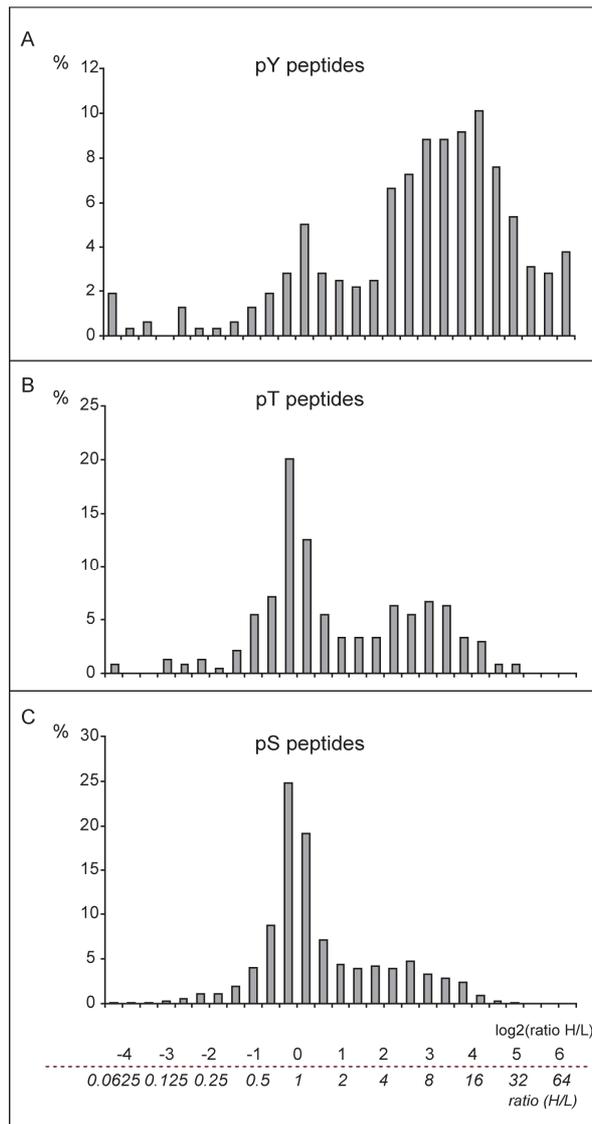


Figure 4.1.16 Phosphatase inhibitors displayed specificity. Percentage of phosphopeptides as a function of up-regulation due to inhibitors. (A) Most of the phosphotyrosine peptides are regulated several fold due to pervanadate treatment. (B) and (C), A proportion, but not all, phosphothreonine and phosphoserine peptides are affected by the treatment of Calyculine A and deltamethrine.

4.1.4 Estimating an upper bound on the stoichiometry of phosphorylation

Quantitative phosphoproteomics, as done with SILAC, compares the relative levels of phosphorylation between two cellular states. However, it is frequently also important to know the degree of phosphorylation or ‘occupancy’ of a given phosphorylation site. There are few available strategies to measure occupancy, particularly in large-scale experiments and they

usually involve the use of isotope labeled synthetic peptides¹⁹³ or require the observation of both the unphosphorylated and the phosphorylated form of the peptide¹⁹⁴.

The work of Erba et.al¹⁹⁵ addressed the concept of using a phosphatase inhibitor to detect phosphorylation occupancy. Several tyrosine phosphorylation sites on EGFR upon EGF stimulation were quantified. By using sodium-pervanadate-induced super-physiological phosphorylation as a reference, the occupancies of the EGF-induced phosphorylation of these tyrosine sites were calculated. However, the endogenous phosphorylation state of these tyrosine sites were not considered, therefore one cannot necessarily conclude that these occupancy rates were induced by EGF treatment. For complete information one also has to include the basic phosphorylation state.

Therefore we chose to include an additional SILAC state. Three cell populations are differentially labeled by SILAC. The light labeled cells serve as control, the medium labeled cells are treated with the stimulus of interest, e.g. a growth factor, and the heavy labeled cells are treated with the phosphatase inhibitor cocktail. The expected pattern of regulated phosphosites is less phosphorylated in the control cells and more phosphorylated in the phosphatase inhibitor treated cells, with the growth factor simulated cells somewhere in between. Importantly, the stoichiometry of phosphorylation after stimulation cannot be higher than the ratio between the heavy and medium labeled forms of the phosphopeptide.

Insulin-induced phosphorylation occupancy

We investigated the phosphorylation occupancy upon insulin stimulation in these mouse liver cells (Figure 4.1.17). We treated triply encoded SILAC liver cells with no stimulus, 5 minutes stimulation with 100 nM insulin, and phosphatase inhibitors (section 3.1.5). In order to obtain a larger number of pTyr sites, we employed a separate tyrosine enrichment step in our experimental protocol. After immunoprecipitation with the pTyr specific antibody 4G10, we quantified a total of 480 pTyr sites in this liver cell line.

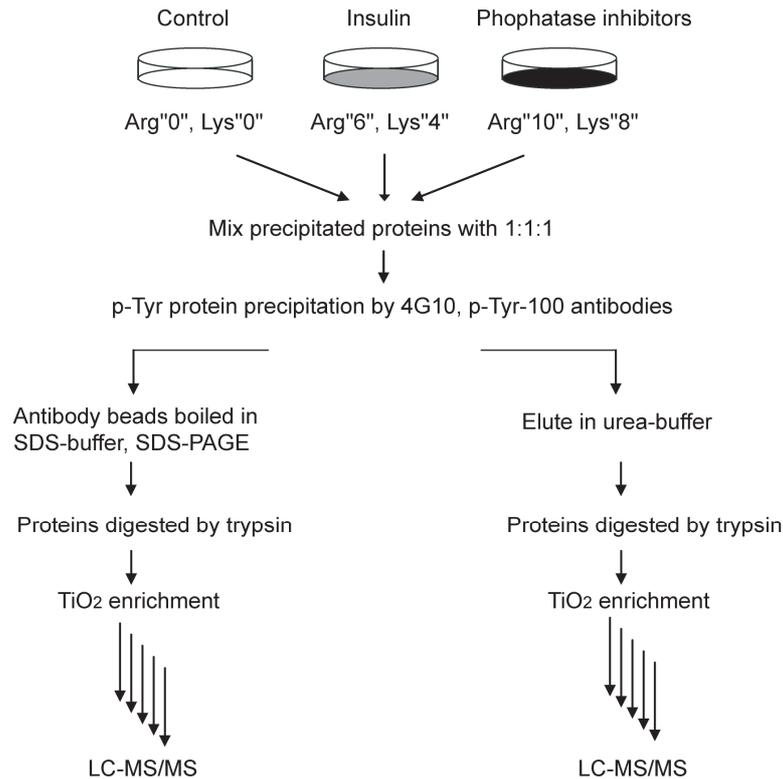


Figure 4.1.17 Experimental design for the phosphorylation site experiment to determine an upper limit of stoichiometry. (A) Cells are triple SILAC labeled. One population serves as control, one is stimulated with insulin and one is treated with phosphatase inhibitor.

As shown in Figure 4.1.18, insulin induced the phosphorylation site of Y1175 in the activation loop of the mouse insulin receptor by more than ten-fold compared to control cells. The phosphopeptide ratio between insulin stimulated and phosphatase treated cells is 1.75. Since a given phosphorylation site cannot be occupied more than 100%, this data puts an upper limit on this phosphorylation site of 60% after 5 minutes insulin stimulation. Figure 4.1.18 also shows the example of a serine phosphorylation site, which is up-regulated compared to basal condition by a factor 2.4. From this it follows that occupancy cannot be greater than 24% under the given condition.

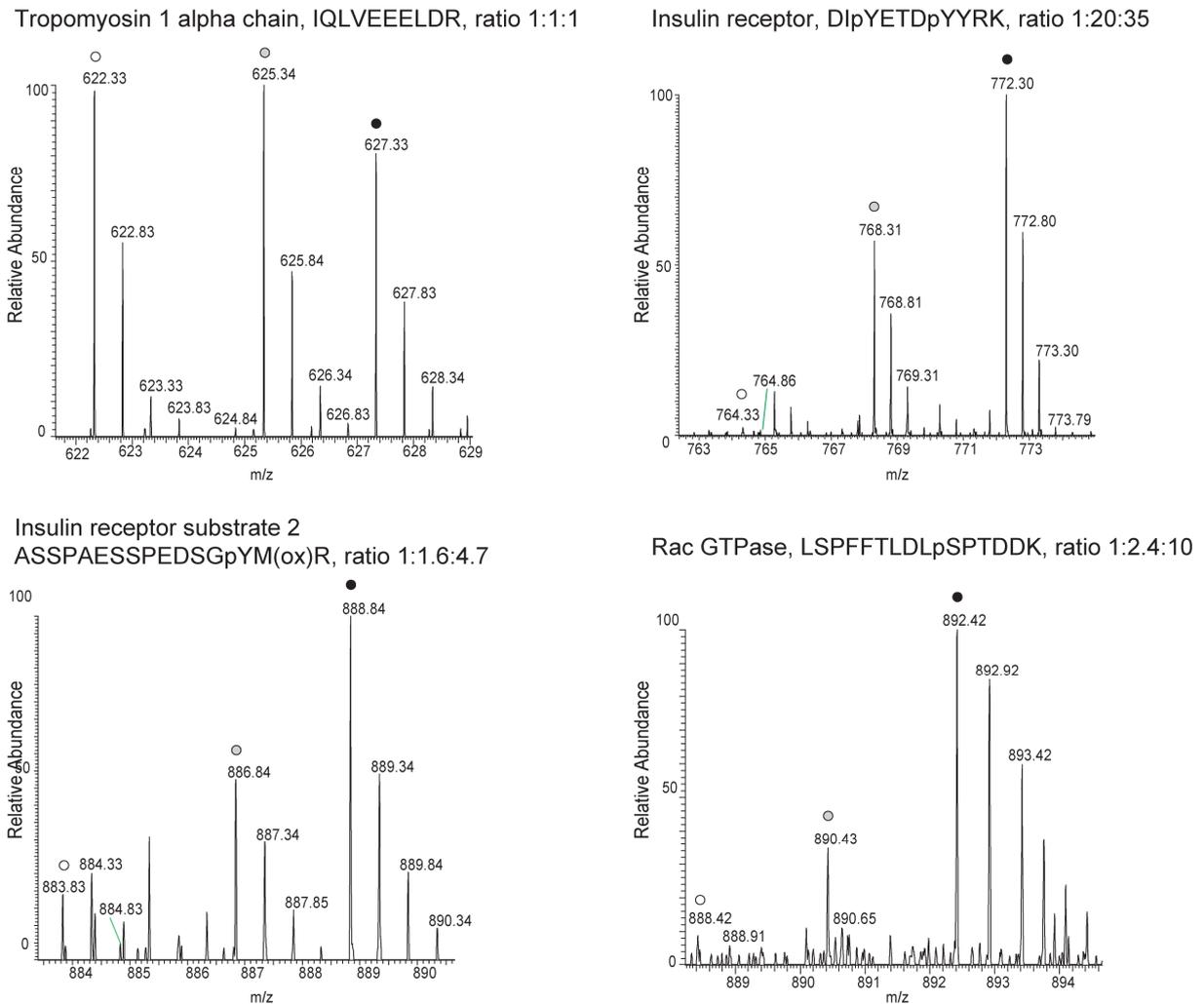


Figure 4.1.18 Examples of peptides with unchanging phosphoprofile and peptides that are phosphorylated in response to insulin and even more increased in response to phosphatase inhibitors.

We have found this method to work more generally for tyrosine phosphorylation than for serine/threonine phosphorylation, because the latter are less induced by the inhibitors (section 4.1.3.2).

4.1.5 Conclusions and discussions

In this project we employed a mixture of three broadband phosphatase inhibitor compounds to boost phosphorylation in live cells. This approach enabled us to determine the phosphoproteome of mouse liver cell line Hepa1-6 in great depth. Phosphorylation is essential

for many cellular processes and malfunction of kinases and phosphatases are often found in cancer. There is no doubt that knowledge of complete phosphoproteomes will greatly facilitate our understanding of fundamental biological principles as well as disease developments at a systematic level. It may even provide guidance to developing therapies and drugs.

In our study, a large Hepal-6 phosphoproteome containing 4,253 accurately assigned phosphorylation sites was derived. General features of the phosphoproteome are similar to what our laboratory reported in a recent large-scale study of human HeLa cells³¹. Interestingly, the phosphoproteome is highly enriched for protein binding functions and interaction domains. This is readily understandable in terms of the signaling functions of this modification. However, this finding also implies that a significant proportion of the phosphoproteome measured here is indeed functional rather than being phosphorylated through an ‘innocent bystander’ effect. If this was not the case, we would have expected preferential ‘background’ phosphorylation of the most abundant proteins, i.e. metabolic enzymes and the like.

For the first time in the phosphoproteomic field, we have assessed the dynamic range of phosphopeptide detection, which turned out to be up to four orders of magnitude with our current technology.

We found quite good overlap with phosphorylation sites recorded in SwissProt, but more than half of the sites found here are nevertheless novel. This suggests that phosphoproteomic is not close to ‘saturation’ yet.

We employed SILAC for quantitation of the phosphoproteome. Here we used the SILAC technology to quantify basal phosphorylation against the degree of phosphorylation after applying a cocktail of phosphatase inhibitors. For phosphotyrosine, inhibition was effective and the majority of pTyr sites were strongly increased upon treatment. For phosphothreonine, and even more for phosphoserine, a majority of sites was unaffected by the inhibitors. This was not due to unusually low doses as our protocol is within the range commonly used in the literature and as cells began to detach after phosphatase treatment. There are several possibilities that could explain the lack of universal up-regulation of pThr/pSer sites. The inhibitors could be specific for only some classes of phosphatases and leave others active. Alternatively, cellular kinase could have a small rate of phosphorylation for a majority of phosphorylation sites. Thus inactivation of phosphatases would not be noticeable for those sites. It would be interesting to follow up on the observations made here by testing panels of

phosphatase inhibitors separately, at different doses and for different times. Regardless of the outcome of such studies, we conclude that broad action and non-specific of serine/threonine phosphatase inhibitors cannot be taken for granted.

The methods applied here are completely generic and could also be used in the same way for testing the specificity of phosphatase inhibitors in development as drugs.

We also introduced several novel analytical concepts in this project. First, we employed phosphatase inhibitors to boost low level phosphorylation sites and make them more likely to be sequenced and identified. Importantly, the presence of the SILAC partner peptide from the unstimulated cells ensures that the phosphorylation site in question is endogenous and not an artifact of the inhibitor. Secondly, we devised a method to put an upper bound on the stoichiometry of sites that are heavily regulated by phosphatase inhibitor relative to control (as shown for insulin regulated sites). As demonstrated here, phosphoproteomics is fast becoming a very streamlined technology that can be used in many different ways to characterize signaling pathways and help in drug development.

4.2 Project 2 - Global effects of kinase inhibitors on signaling networks revealed by quantitative phosphoproteomics

This work is included in a manuscript in preparation:

Global effects of kinase inhibitors on signaling networks revealed by quantitative phosphoproteomics

Cuiping Pan et al.

Phosphorylation regulates a vast number of cellular processes and is of pivotal importance in signal transduction. Abnormal phosphorylation is found in various diseases, especially in cancers, where kinase aberration (usually secondary to genetic alterations) is often critical for their etiology and pathology⁵⁴. For this reason much effort has been devoted to understanding kinase function and to manipulating their actions. One widely adopted strategy is to suppress kinase activity by small molecule inhibitors. While chemical inhibitors have been used in laboratory research for decades, the concept of specific kinase inhibitor drugs was only established in recent years. Today, kinase inhibitor compounds amount to about 30% of drug development programs in the pharmaceutical industry¹⁹⁶.

Techniques to measure inhibitor effects were only established in a high throughput manner in the last ten to twenty years. Conventionally interactions between the inhibitor compound and a panel of selected kinases are profiled *in vitro*^{71, 197}. When systematic screening methods are introduced, these enzyme assays can include more kinases and cover the kinome more broadly. As a result, tests that are less biased can be performed. Examples of these advanced methods include phage plaque assays combined with quantitative PCR¹⁹⁸, yeast three-hybrid¹⁹⁹, and mass spectrometry with affinity chromatography^{200, 201}. While these *in vitro* studies greatly help in defining the targets of kinase inhibitors, they have several limitations. Chemical or genetic modifications are often required, such as fusing kinases with bacterial phage proteins in phage plaque assays, with linker proteins in the yeast three-hybrid approach or addition of

chemical linkers to inhibitor compounds in affinity chromatography purification. These manipulations are often laborious and time-consuming. Besides they risk changing the properties of kinases and inhibitor compounds. Furthermore, these methods can only address the direct targets of the inhibitor compounds but not their influence on the cellular signaling network in total. For example, the same kinases can be employed by multiple signaling pathways, thus their functions on downstream signaling events can be difficult to predict. This complexity complicates the picture and therefore it is difficult to estimate or predict the overall effects of inhibitor compounds in a cellular environment. Certainly a cell-based approach, which allows system-wide elucidation of the effects of kinase inhibitors, will further improve the target evaluation process.

While cell imaging provides direct visualization of the kinase inhibition effects *in vivo*²⁰²⁻²⁰⁵, cell system based quantitative mass spectrometry is able to simultaneously monitor protein expression and modification events at a much broader scale. Stable-isotope labeling by amino acids in cell culture (SILAC), a metabolic labeling method applied during cell growth, generates completely stable isotope labeled cell populations that are otherwise equal to non-labeled cells. This system enables a direct comparison of several cell populations with different biological or chemical treatments. For example, when SILAC was applied to study the effect of the Her2 kinase inhibitor, cellular changes of tyrosine phosphorylated proteins due to this treatment could be quantified²⁰⁶.

In recent years, studies of phosphorylation have been greatly enhanced by progresses in mass spectrometric techniques. Significantly, phosphopeptide enriching methods such as IMAC and titanium sphere chromatography have demonstrated detection of thousands of phosphorylation sites, completely changing the capabilities of the phosphoproteomics field^{28, 94-97}. We reasoned that the combination of SILAC and the most recent phosphoproteomics technique should provide a powerful tool to explore the effects of kinase inhibitors on the overall cellular signaling network. This could help translating our knowledge of kinase inhibitor effects from the unit of individual proteins to the level of the whole (phospho)-proteome. Furthermore, modern phosphoproteomics makes it possible to identify and quantify individual phosphorylation sites. As multiple phosphorylation sites on the same protein can have different functions and can be regulated differently, this detailed information can precisely pinpoint the responsive sites and provide a much more detailed picture of the inhibitor effects.

As proof of principle, we first examined two widely used inhibitor compounds in signal transduction laboratories: U0126 as a MEK1/2 inhibitor and SB202190 as a p38 α / β inhibitor. Next, we applied the same technique to screen the effects of dasatinib, a clinical drug used to inhibit mutated BCR-ABL in chronic myelogenous leukemia (CML).

4.2.1 Effects of U0126 and SB202190 on the EGFR signaling pathway

U0126 and SB202190 are very potent, cell-permeable inhibitors for MEK1/2 and p38 α / β , respectively. Their corresponding IC₅₀ values are at the sub-micromole range. Previous studies characterized both as highly selective inhibitor compounds^{71, 198}. However due to their pharmacological limitations, neither of the two inhibitors entered clinical trials but they are instead used as research compounds in laboratories. Thousands of publications employed these compounds and various kinase functions have been derived.

SB202190 is a pyridinyl imidazole (Figure 4.2.1). It binds to the ATP pocket of p38 α / β , and therefore precludes ATP binding to the kinase. Since ATP pockets are conserved among kinases, it is suggested that the specificity of the inhibitor compound is largely achieved by utilizing other residues than the direct ATP binding sites. A likely additional interaction site on the inhibitor is the 4-phenyl ring^{54, 207}.

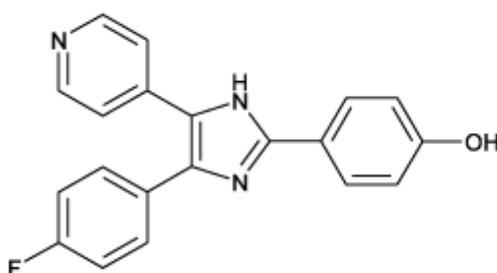


Figure 4.2.1 Chemical structure of SB202190, a p38 α / β MAPK inhibitor. Its chemical name is 4-(4-Fluorophenyl)-2-(4-hydroxyphenyl)-5-(4-pyridyl)-1H-imidazole, with the molecular formula: C₂₀H₁₄FN₃O. Molecular mass is 331.34 g/mol.

Different from SB202190 and most other inhibitor compounds, U0126 (Figure 4.2.2) does not compete with ATP. Therefore the high ATP concentration in the intracellular milieu (2-10

mM) has only minimum influence and U0126 represents a unique group of kinase inhibitors. U0126 exerts its inhibitory effect mainly by binding to inactivated MEK1/2 and precludes allosteric change of the kinase domain. Therefore MEK1/2 are locked in their inactive confirmation⁷¹.

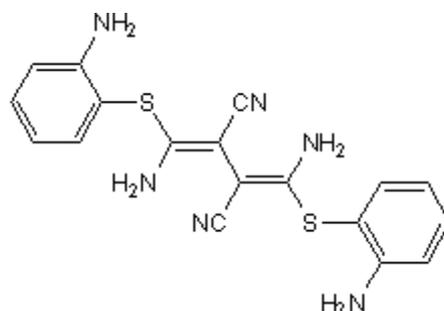


Figure 4.2.2 Chemical structure of U0126, a MEK1/2 inhibitor. Its chemical name is 1,4-diamino-2,3-dicyano-1,4-bis[2-aminophenylthio] butadiene, with the molecular formula: $C_{18}H_{16}N_6S_2$. Molecular mass is 380.50 g/mol.

MAPK family proteins propagate signals for various growth factors and inflammatory stimuli. MAPK signaling consists of several tiers: upstream signaling proteins such as growth factor receptors, MAP3Ks, MAP2Ks, MAPKs, and downstream proteins such as the ribosomal protein S6 kinase p90RSK. There are four major MAPK cascades: ERK1/2, ERK5, JNKs, and p38 MAPKs. These cascades are similar but can perform different cellular functions^{208, 209}. To investigate their specific kinase functions it is important to have highly selective kinase inhibitor compounds. For example, to study ERK1/2-dependent cellular functions, U0126 and PD 98059 are often employed as ERK1/2 are generally considered as the highly dominant, if not the only, physiological target of MEK1/2 (also called MAP2K1/2). For elucidating the cellular functions of p38 α/β , inhibitor compounds such as SB202190 and SB203580 are frequently applied.

ERK1/2 and p38 α/β can be activated by EGF in HeLa cells³¹. Based on this cell system, we examined the effect of SB202190 and U0126 on the overall signaling network. We set up three conditions in which HeLa cells were SILAC-labeled with distinct forms of arginine and lysine. Light labeled cells (Arg0, Lys0) remained untreated and served as control. Medium labeled cells (Arg6, Lys4) were stimulated with growth factor, whereas heavy labeled cells (Arg10, Lys8) were treated with both kinase inhibitor and growth factor stimulation (Figure 4.2.3 A).

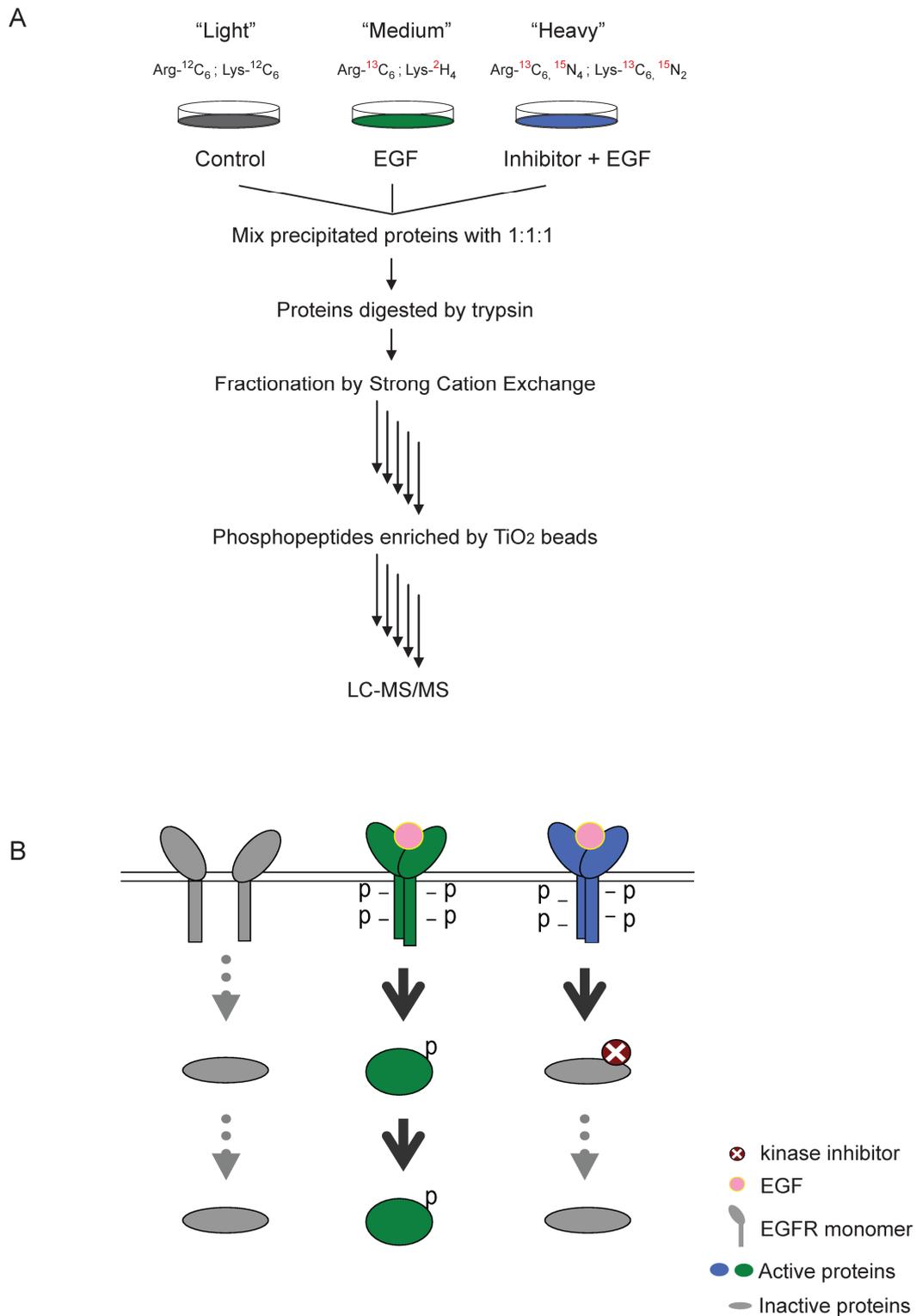


Figure 4.2.3 Strategy to detect the effects of kinase inhibitors on cellular signaling network. (A) Three SILAC-labeled cell populations are subjected to different treatments as indicated. (B) Principle response of the targeted kinase and its downstream substrates in the three SILAC conditions.

For the two inhibitor treatments a concentration of 10 μM was applied to heavy labeled cells for 20 minutes. In the presence of the inhibitor we applied 150 ng/ml of EGF to the cells for another 15 minutes. The medium labeled cells were treated with 150 ng/ml of EGF for 15 minutes. Combined cell lysates from these three conditions were subjected to our phosphoproteomics processing workflow as described in section 3.3. A simple diagram of the targeted kinase and their downstream substrates is displayed in Figure 4.2.3 B.

4.2.1.1 Identification and quantitation results

Mass spectrometric data from the two inhibitor experiments were processed and quantified with in-house developed software MaxQuant. Identification was performed by searching the IPI_human protein database using the Mascot algorithm. Statistical calculations are based on combined Mascot plus post-translational modification (PTM) scores. The following criteria have to be met for all included peptides: peptide posterior error probability (PEP) ≤ 0.1 , and peptide false discovery rate (FDR) ≤ 0.01 (see section 3.4). Using the PTM score³¹, we could localize the phosphate groups with high confidence (Class I phosphosites) in 5,497 cases. A total of 5,748 phosphopeptides were quantified from 2,048 proteins.

Quantitation results of experiments with both inhibitors are listed in Table 4.2.1. Phosphopeptides were grouped into three major categories based on their response to EGF, namely up-regulated (Ratio_{ML} ≥ 2), down-regulated (Ratio_{ML} ≤ 0.5), and non-changed ($0.5 < \text{Ratio}_{ML} < 2$). Within each category, phosphopeptides were further sorted according to their response to the kinase inhibitor. The majority of the phosphopeptides (~85%) were influenced by neither the growth factor nor the kinase inhibitors. Even though the overall numbers of quantified phosphopeptides are different in both experiments, the percentages of EGF up- and down-regulated phosphopeptides are similar, with ~10% upregulated and ~2% downregulated in each case. This indicates that the difference in overall phosphopeptide numbers is mainly due to technical factors.

Not surprisingly, multiple phosphopeptides of the same protein can have different response patterns. A striking example is the neuroblast differentiation-associated protein AHNAK. Among 50 quantified phosphopeptides of AHNAK, six of the nine possible response patterns were found.

Table 4.2.1 Quantitation results of the U0126 and SB202190 experiments.

| | M/L \geq 2.0 | | | M/L \leq 0.5 | | | 0.5<M/L<2.0 | | |
|--|----------------|----|-----|----------------|----|----|-------------|----|----|
| Pattern | P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 |
| | | | | | | | | | |
| U0126 (MEK-inh) | 162 | 10 | 151 | 43 | 17 | 3 | 3102 | 23 | 67 |
| SB202190 (p38 α / β -inh) | 205 | 1 | 277 | 50 | 38 | 4 | 4122 | 31 | 94 |
| Overlap | 100 | 0 | 78 | 7 | 7 | 0 | 2528 | 4 | 8 |

M/L represents ratio of “medium labeled cells” (EGF stimulation) versus “light labeled cells” (inhibitor pretreatment and EGF stimulation). Within each category, ratios are further classified according to their responses to the inhibitor. Up-regulation: Ratio \geq 2.0; down-regulation: Ratio \leq 0.5.

4.2.1.2 Effects of the inhibitors on MEK1/2 and p38 α / β signaling branches

Known and novel substrates of MEK1/2 and p38 α / β

Phosphorylation of the activation loop directly reflects kinase activity. We captured the relevant phosphopeptides from the activation loops of ERK1/2 and p38 α in both experiments. In agreement with many other reports, they were highly induced by EGF but suppressed by their corresponding inhibitors without cross reaction (Figure 4.2.4). Several specific downstream targets of these inhibited kinases, such as Thr69/Thr71 in the transcription factor ATF2²¹⁰ and Thr359/Ser363 in the kinase p90RSK²¹¹, displayed the same response patterns. Some phosphorylation sites of the upstream proteins were not affected. Examples include Ser991/Ser995 of the EGFR.

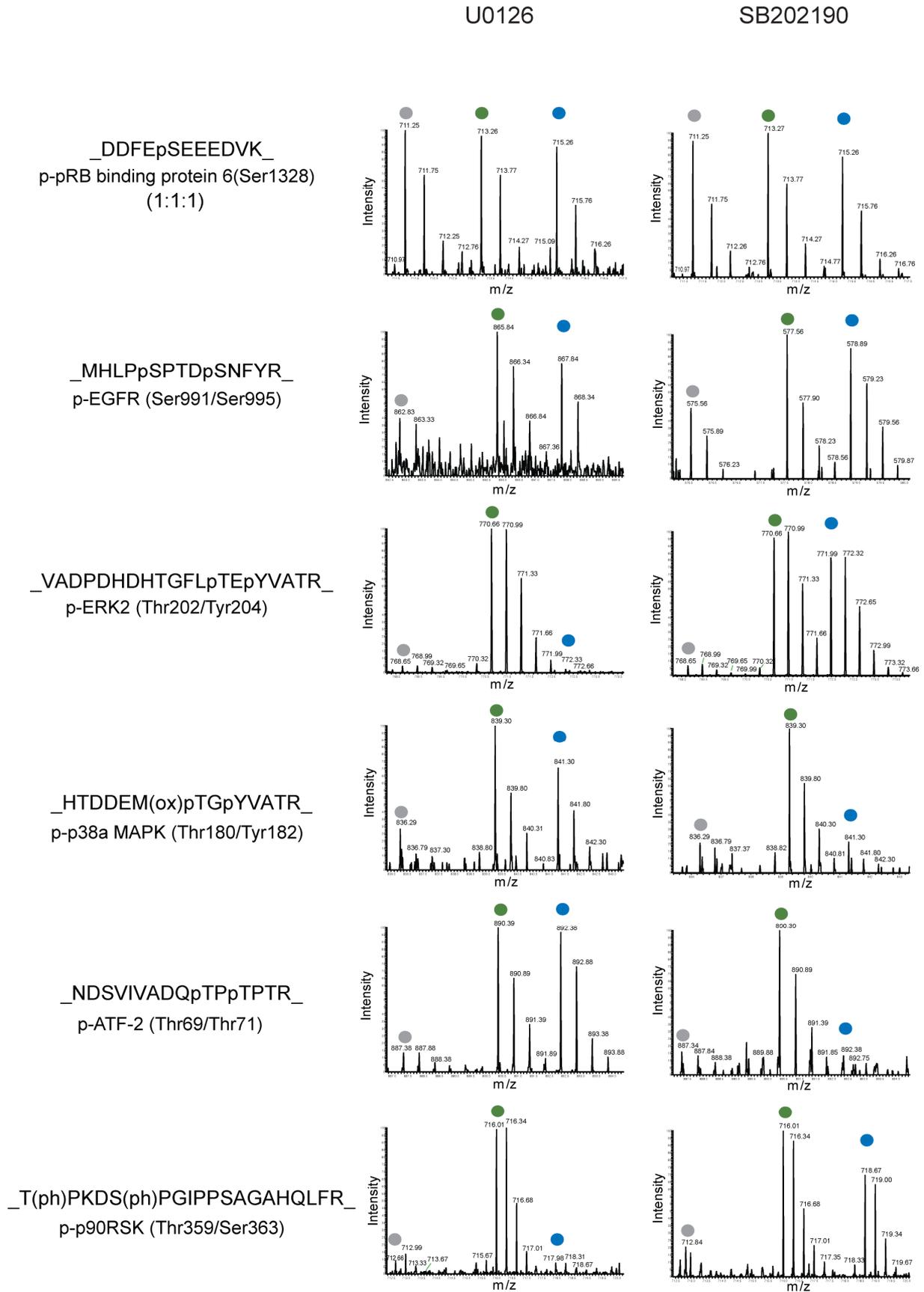


Figure 4.2.4 Effects of the inhibitors on known substrates of signaling branches of ERK1/2 and p38 α/β .

In principle, downstream substrates of the targeted kinase should display the same response pattern as the kinase itself. However, a different response pattern can still be observed for the kinase downstream substrates if a phosphatase is involved or the signaling kinetics at later stages is considered (for example, the P9 pattern in Table 4.2.1). Phosphatases reverse phosphorylation and therefore decrease the intensity of phosphopeptides in MS. Under the specific EGF stimulation condition in this study, those late regulation events of phosphorylation may not have happened. To simplify further analysis, we will only consider those phosphopeptides with the same response pattern to the kinase (up upon EGF stimulation and down upon inhibitor treatment; P3 in Table 4.2.1) as potential direct or indirect downstream substrates. On this basis, there are 151 potential peptide substrates of MEK1/2, and 276 (excluding p38 α/β) potential peptide substrates of p38 α/β . There are 78 shared peptides between these two groups of potential kinase substrates. Analysis in Gene Ontology indicates that their corresponding proteins are involved in diverse biological processes but their molecular function is dominantly ‘binding’ (e.g. protein binding, nucleotide binding, and metal ion binding).

Crosstalk between MEK1/2 and p38 α/β signaling pathways

To better define shared and specific substrates, we selected those peptides that were quantified in both inhibitor experiments. Again to simplify interpretation, EGF up-regulation ($\text{Ratio}_{\text{M/L}} \geq 2$) was set as a prerequisite. Phosphosites with accurate localization, i.e. class I sites, are classified into three groups (Figure 4.2.5 A): MEK1/2 specific (“U” group), p38 α/β specific (“S” group), and shared (“X” group) phosphosites.

Proline at the +1 position following pSer/pThr is a well accepted kinase motif for MAPKs. Within each group we analyzed the adjacent amino acid residues surrounding each phosphorylation site. Interestingly, the percentages of proline-directed phosphopeptides (pS/pT-P motif) are low, ranging from 15% to 30% (Figure 4.2.5 B). This indicates that the majority of these potential substrates are secondary or further downstream substrates of MEK1/2 and p38 α/β . This is not unexpected, as MAP2Ks and MAPKs are generally located in a relatively upstream position in growth factor signaling pathways.

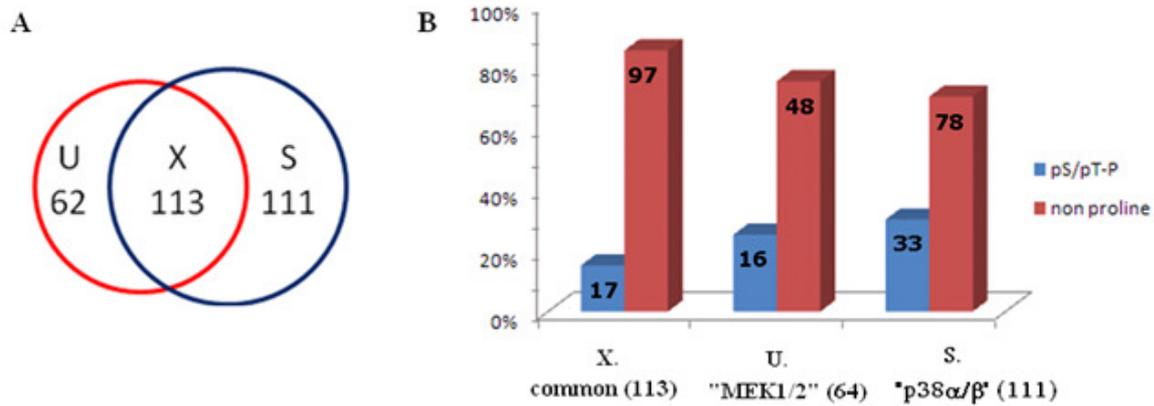


Figure 4.2.5 (A) Overlap between potential phosphosite substrates of MEK1/2 (Red circle) and p38α/β (blue circle). (B) Distribution of proline-directed and non-proline-directed phosphorylation sites. U (U0126) group - specific to MEK1/2, S (SB202190) group - specific to p38α/β, and X group - shared.

Although ERK1/2 and p38α/β share 50% sequence similarity, they are employed in different cellular responses²⁰⁹. ERK1/2 are activated by various stimuli to regulate meiosis, mitosis, and postmitotic functions in differentiated cells. p38 MAPKs are strongly activated by cytokines and stress factors, including osmotic shock and heat shock, and to regulate immune and stress responses. The conserved motif in kinase activation loop is Thr-Glu-Tyr (TEY) for ERK1/2 and Thr-Gly-Tyr (TGY) for p38α/β. Previously, comparison of the substrates for ERK1/2 and p38α/β was reported at the transcription level²¹². Here we are able to compare their substrates at the phosphorylation level in the overall signaling network. Generally, phosphorylation involves rapid and short-term regulation events whereas transcription starts later, typically after 30 minutes. Our comparison is based on an active signaling network after 15 minutes of EGF stimulation, where the signal was propagated to various downstream effectors and the forward signal dominated over the feedback signal. It is within this network that we discovered many differentially regulated phosphosites by MEK1/2 and p38α/β. These phosphosites come from important signaling proteins, such as JNK3, PLC, MSK2, HSP27, death-inducer obliterator 1, DNA topoisomerase 2-alpha, and so on.

Requiring EGF up-regulation as a prerequisite, we have limited our analysis to a small, more easily analyzable dataset. Still, caution should be exercised in data interpretation. U0126 was previously reported to inhibit ERK5, a fourth family member of MAPKs besides ERK1/2, JNKs, and p38 MAPKs. ERK5 shares the same consensus motif TEY as ERK1/2. However,

ERK5 is unique in its carboxy-terminal half and possesses a nuclear translocation signal in its sequence. It is involved in cardiovascular development and neural differentiation²⁰⁸. Due to their distinctive features, the downstream substrates of ERK5 and ERK1/2 may be different. Suppressing both at the same time can cause ambiguity in assigning the structure of the network. From the perspective of drug effect screening, it is precisely the strength of our unbiased, systematic approach using quantitative MS that it provides a clear view of the overall effects of the inhibitor treatment.

Evaluation of EGF signal propagation

Based on the analysis of phosphorylation sites, we created a rough model of EGF signal propagation upon 15 minutes of 150ng/ml EGF stimulation (Figure 4.2.6).

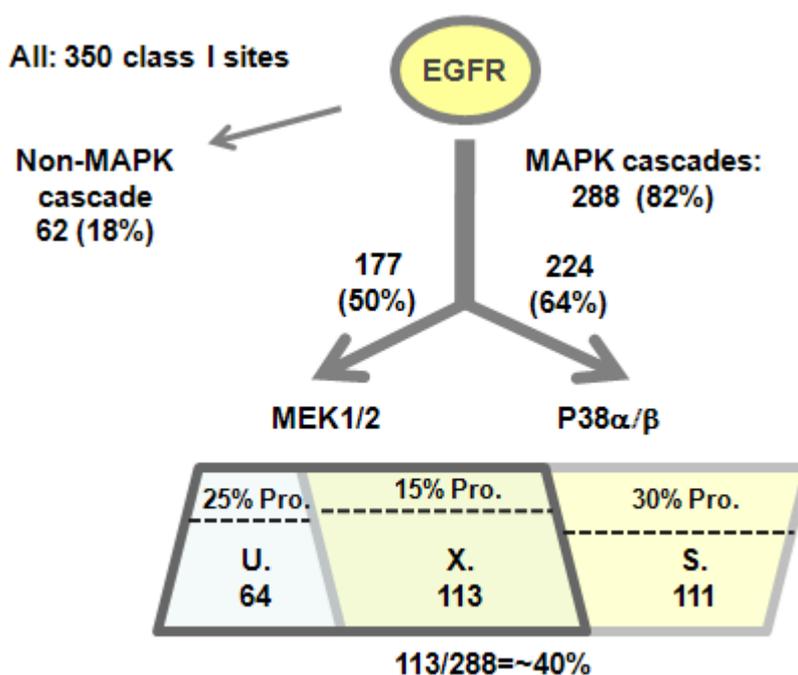


Figure 4.2.6 A simple model of EGF signal propagation. Overall 350 class I phosphosites were up-regulated by EGF in both inhibitor experiments. The percentages of proline-directed phosphorylation events are listed. The three groups (U, X, and S) are the same as in Figure 4.2.5.

Overall 350 class I phosphorylation sites were activated by EGF. Of these, 82% participated in MAPK cascades as judged by their suppression after MAPK inhibitor treatments, leaving only a small role for other signaling cascades. Within the MAPK cascades, the signal diverged to MEK1/2, p38 α / β and other branches. The fraction of proline-directed peptide substrates is only 15 – 30%, indicating that the signal spreads from a limited set of direct substrates to a broad set of secondary and further downstream effectors. Although only a rough estimate, this model is very interesting in itself. It suggests that MAPK cascades are predominant in these growth factor induced signal transduction events. Again, it clearly demonstrates the strength of quantitative MS and systems biology approach in reconstructing the signaling network.

4.2.1.3 Potential side effects of the inhibitor compounds

Off-target effects are a major concern for inhibitor compound development. As most kinase inhibitors are ATP competitors and ATP binding pockets are conserved among kinases, it is likely that inhibitor compounds suppress more kinases than the targeted one. The fact that these suppressed kinases can have limited sequence homology further complicates analysis. Besides the off-target issue, there is an equally important question in the field, namely the secondary effect of the kinase inhibitor compound on the signaling pathway. This issue has been addressed to a lesser degree and in a less systematic manner due to the technical limitations. The strength of our approach is to examine the global effects of kinase inhibitors over the entire network.

Previously we discussed the potential peptide substrates of MEK1/2 and p38 α / β . Here we set out to inspect some other phosphorylation changes that can potentially introduce drastic effects into the signaling network. Phosphopeptides that were up-regulated by the inhibitor compounds compared to EGF only treatment ($\text{Ratio}_{H/M} \geq 2$) are at first unexpected and demonstrate the ability to distort EGF effects and therefore are of great interests.

In the SB202190 dataset, 70 phosphopeptides fulfilled this criterion. Among the list are Tyr301 of A-Raf and three phosphosites (Thr58, Ser61, and Ser63) from PP2A regulatory subunit beta. A-Raf closely resembles Raf-1, a major signaling molecule of the EGF response, in sequence and activation mechanism. Ras-GTPase and Src can activate both Rafs by phosphorylating two adjacent tyrosine residues, Tyr-301/302 in A-Raf and Tyr-340/341 in Raf-1²¹³. We detected reduced phosphorylation of Tyr301 in A-Raf upon EGF treatment ($\text{Ratio}_{M/L}=0.5$). However SB202190 treatment reversed this down-regulation ($\text{Ratio}_{H/M}=9.5$)

such that this phosphorylation site is now up-regulated instead. This suggests that the A-Raf signaling branch is greatly activated after the inhibitor treatment. The reversed phosphorylation of the regulatory subunit of phosphatase PP2A ($\text{Ratio}_{M/L}=0.1$, $\text{Ratio}_{H/M}=8$) likewise could induce a drastic change, given the fact that PP2A governs dephosphorylation events on various signaling molecules.

Together, 50 phosphopeptides in the U0126 dataset display $\text{Ratio}_{H/M} \geq 2$. Interesting examples include the activation motif of JNKs (Thr221/Tyr223 with $\text{Ratio}_{M/L}=2.6$, $\text{Ratio}_{H/M}=6.5$), IRS-2 (Ser596 with $\text{Ratio}_{M/L}=1.1$, $\text{Ratio}_{H/M}=2.2$), and HSP27 (Ser15 with $\text{Ratio}_{M/L}=3.6$, $\text{Ratio}_{H/M}=6.9$ and Ser26 with $\text{Ratio}_{M/L}=5.2$, $\text{Ratio}_{H/M}=12$). JNKs are well-known stress-activated protein kinases and constitute an important branch of MAPK cascades. HSP27 is phosphorylated at Ser15, Ser78 and Ser82 by MAPKAP kinase 2 as a result of activation of the p38 MAP kinase pathway²¹⁴. The further up-regulation of JNK pathway and p38 MAPK pathway by U0126 treatment indicates a potential interplay among several MAPK cascades. As distinctive functions of these MAPK family proteins have been reported, attention should be paid to the effect of the kinase inhibitors on them.

Importantly, proteins regulating transcriptional process are found in both datasets with $\text{Ratio}_{H/M} \geq 2$ from SB202190 and U0126 experiments. It indicates a long-term effect of the inhibitors at the transcriptional and protein expression levels.

4.2.2 Effect of dasatinib on the BCR-ABL signaling pathway

Chronic myelogenous leukemia (CML) is characterized by over-proliferation of myeloid cells. The fundamental principle of its etiology is the fusion of two chromosomes to produce the chimeric protein BCR-ABL. In normal physiology Bcr is encoded by chromosome 22 whereas the non-receptor tyrosine kinase ABL is encoded by chromosome 9. In CML, however, the long arms of chromosome 22 and chromosome 9 are fused together, resulting in the so-called Philadelphia chromosome and the constitutively activated tyrosine kinase BCR-ABL²¹⁵. Two predominant BCR-ABL fusions are involved in CML²¹⁶: fusion of the first 13 exons of BCR to exon 2 of ABL results in the b2a2 fusion (e13a2), and fusion of the first 14 exons of BCR to exon 2 of ABL results in the b3a2 fusion (e14a2).

Treatment of CML has been greatly advanced by using small inhibitor compounds to selectively inhibit the kinase activity of BCR-ABL. The first BCR-ABL inhibitor drug Gleevec[®], or imatinib, was rapidly approved by Food and Drug Administration (FDA) in the US in the beginning of the 21st century. In fact, the success of imatinib for the first time proved the novel concept of using small kinase inhibitor compounds as rational drugs and opened a new field in drug development. This achievement was so striking that an editorial essay from *The Oncologist* praised it as breaking the barrier of “The oncologic four-minute mile”²¹⁷. Later, a second generation of inhibitor drugs appeared in the clinic, with the aim of inhibiting point-mutated versions of BCR-ABL. Among these is dasatinib (Figure 4.2.7), a highly potent, orally active inhibitor for both inactive and active BCR-ABL. Dasatinib is most potent towards chronic phase CML, with complete hematologic responses in 90% of patients, 52% of whom achieved a major hematologic response²¹⁸. Distinct in chemical structure from imatinib, dasatinib is able to inhibit most BCR-ABL variants found in CML patients²¹⁹. In cell line tests, dasatinib displayed an over 300 fold higher potency than imatinib²²⁰. Today the combination of imatinib and dasatinib is recommended for CML therapy.

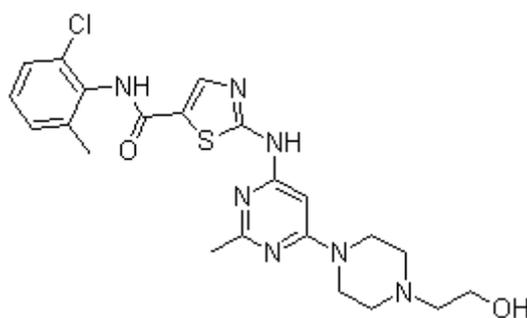


Figure 4.2.7 Chemical structure of dasatinib. Its chemical name is N-(2-Chloro-6-methylphenyl)-2-(6-(4-(2-hydroxyethyl)-piperazin-1-yl)-2-methylpyrimidin-4-ylamino)thiazole-5-carboxamide, with the molecular formula: C₁₈H₁₆N₆S₂. Molecular mass is 380.50 g/mol.

Dasatinib binds to the kinase domains of ABL kinase (Figure 4.2.8). It has a similar potency towards Src family kinases and PDGFR²²¹. Research on dasatinib's mechanism of action focuses on two major themes: the direct binding targets and more downstream signaling molecules. Mass spectrometry has played an important role in these investigations. Recently Goss et al²¹⁶ analyzed immunoprecipitated protein in tandem MS and derived a common phosphotyrosine signature for BCR-ABL in six different CML cell lines. Hantschel *et al*²²²

and Bantscheff *et al*²⁰¹ combined affinity purification technique with quantitative mass spectrometry to screen the binding targets of dasatinib. In the study from Bantscheff *et al* seven broad-band inhibitors were immobilized in one affinity column (kinomebead) to test the competition of certain kinase inhibitor compounds with those unspecific ligands. Interestingly these kinase beads are able to cover 80% of the phylogenetic tree.

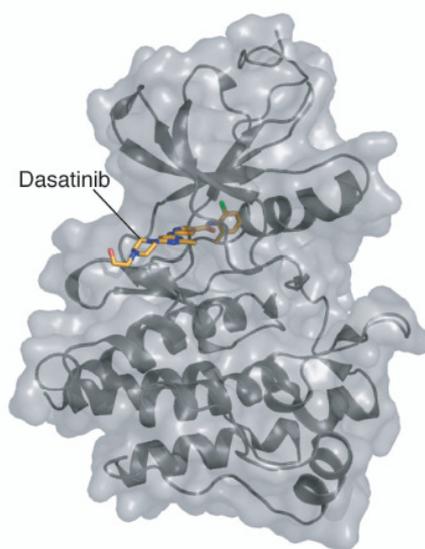


Figure 4.2.8 Crystal structure of the Abl kinase domain complexed with dasatinib. From Hantschel *et al.*²²², based on *PDB ID code 2GQG*

Here we describe a simplified assay to elucidate the effects of this drug on signaling pathways. It employs SILAC-based phosphoproteomics to analyze, for the first time, the effect of a current cancer drug in a cell line without any bias and in a systems-wide manner.

The human immortalized myelogenous leukemia cell line K562 bears the e14a2 fusion and expresses constitutively activated BCR-ABL. Therefore it serves as a suitable cell model for studying CML. In our experiments, three groups of K562 cells were cultured in “light”, “medium” and “heavy” SILAC conditions (Figure 4.2.9). In contrast to the MAPK inhibitor studies described above no stimulus is necessary because BCR-ABL itself provides a constitutive signal. While “light” cells were treated with DMSO only, “medium” and “heavy” cells were treated with 5 nM and 50 nM dasatinib for one hour, respectively. In this way, a dose dependence can in principle be measured. Cells were harvested and enriched for

phosphopeptides as described in section 3.1.5 and 3.3, respectively. Data analysis was performed according to section 3.4. The resulting dataset has 99% identification confidence.

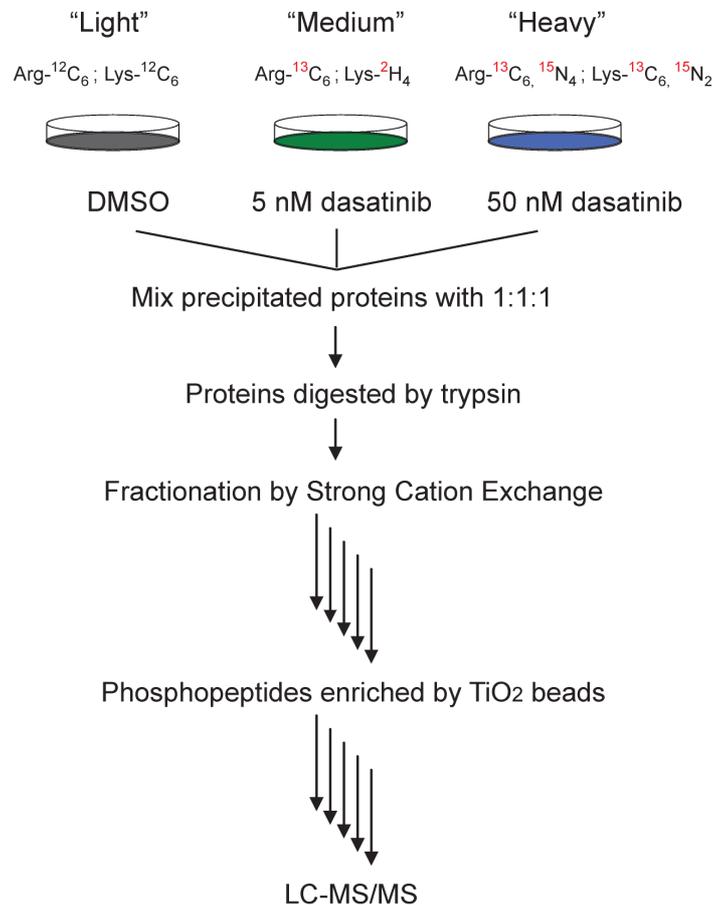


Figure 4.2.9 Strategy to detect the effect of dasatinib on the cellular signaling network. Three populations of SILAC cells were treated with control solution and two different inhibitor concentrations for one hour to examine phosphoproteome changes.

4.2.2.1 A large number of phosphosites are suppressed by dasatinib

We identified 4,833 phosphopeptides from 1,804 proteins, from which 4,286 class I phosphosites were derived. About 12% of these phosphopeptides were suppressed by 5 nM dasatinib treatment, while a ten-times larger dose of dasatinib down-regulated 16% of the phosphopeptides (Table 4.2.2). As expected, there is much more down-regulation than up-regulation. Reassuringly, most regulated phosphopeptides and phosphosites in the treatment with low dose were also quantified as down-regulated sites in the experiment with high dose inhibition (79%). This also implies that these large-scale quantitative phosphorylation

experiments were quite reproducible. Interestingly, there was a small proportion of up-regulated sites as well. However, overlap between these sites is low (~10%). When the criterion of down-regulation is loosened to 50% change (Ratio \geq 1.5), the overlap increases to ~30% (69 phosphopeptides), with 219 down-regulated phosphopeptides in 5 nM dasatinib treatment and 206 in 50 nM dasatinib treatment. These sites will not be discussed further.

Table 4.2.2 Quantitation of phosphopeptides from dasatinib treated K562 cells.

| | | Phosphopeptides | Class I phosphosites |
|---------------------------------------|---------|-----------------|----------------------|
| All | | 4833 | 4286 |
| Down-regulation (Ratio \leq 0.5) | 5 nM | 587 | 524 |
| | 50 nM | 778 | 708 |
| | overlap | 465 | 424 |
| Up-regulation (Ratio \geq 2) | 5 nM | 52 | 45 |
| | 50 nM | 33 | 29 |
| | overlap | 6 | 3 |

Dasatinib is known to down-regulate several important signaling proteins. Table 4.2.3 lists some of the well-known proteins that were captured in our study. The activating autophosphorylation sites of ABL1 (in the fusion protein), SRC, MAPK1 and MAPK3 were significantly suppressed, indicating that their kinase activities were reduced. Additionally, tyrosine phosphorylation of BCR, LYN and BTK was drastically down-regulated. The pThr-58 of c-Myc is known to be regulated by mitogen stimulation. It is crucial for various functions of c-Myc, such as transcription regulation²²³, altered intracellular location²²⁴, and protein degradation²²⁵. Following upstream kinase inhibitions, Thr-58 was largely dephosphorylated after dasatinib treatment in our experiments. This demonstrates that large-scale quantitative phosphoproteomics can pin-point crucial sites involved in cell proliferation.

Table 4.2.3 Marker phosphopeptides downregulated by dasatinib.

| Name | Phospho position | Nr. phospho | Sequence | M/L | H/L |
|-------|------------------|--------------------------------------|---------------------------------------|------|------|
| BCR | 122 | 1:S | _ASASRPQPAPADGADPPPAEEPEARPDGEGpSPGK_ | 0.3 | 0.3 |
| | 644 | 1:Y | _NSLETLLpYKPVDR_ | 0.3 | 0.2 |
| ABL1 | 805, 808 | 2:S,S | _DIMEpSpSPGSSPPLTPK_ | 0.6 | 0.5 |
| | 229 (568) | 1:S | _NKPTVYGvSPNYDK_ | 0.1 | 0.1 |
| | 226 (565) | 1:Y (autophosphorylation site) | _NKPTVpYGVSPNYDK_ | 0.3 | 0.1 |
| SRC | 427;373 | 1:T | _LIEDNEYpTAR_ | 0.4 | 0.04 |
| | 426;372 | 1:Y (autophosphorylation site) | _LIEDNEpYTAR_ | 0.4 | 0.1 |
| MAPK1 | 185 | 1:T | DHTGFLTEYVATR | 0.03 | 0.03 |
| | 185,187 | 2: T,Y (autophosphorylation site) | _VADPDHDHTGFLpTEpYVATR_ | 0.1 | 0.02 |
| MAPK3 | 202,207 | 2: T,T | _IADPEHDHTGFLpTEYVApTR_ | 0.1 | 0.05 |
| | 204 | 2: T,Y (autophosphorylation site) | _IADPEHDHTGFLpTEpYVATR_ | 0.04 | 0.03 |
| MYC | 58 | 1: T | _KFELLPpTPPLSPSR_ | 0.3 | 0.3 |
| BTK | 147, 149 | 2: S,S | _NGpSLKPGpSSHrk_ | 0.5 | 0.3 |
| | 627, 631 | 2: Y,Y | _VpYTIMpYSCWHEKADER_ | 0.03 | 0.3 |
| LYN | 508 (578) | 1: Y | _AEERPTFDYLSVLDFFYTATEGQpYQQQP- | 0.3 | 0.2 |

4.2.2.2 Proline-directed phosphorylation is dominant in BCR-ABL signaling

To evaluate the signaling affected by dasatinib, we selected a sequence window of 13 amino acids with the down-regulated phosphorylation sites positioned in the center. The derived sequences for those 424 commonly down-regulated phosphorylation sites were submitted to WEBLOGO (<http://weblogo.berkeley.edu/logo.cgi>) for assessing kinase substrate motifs.

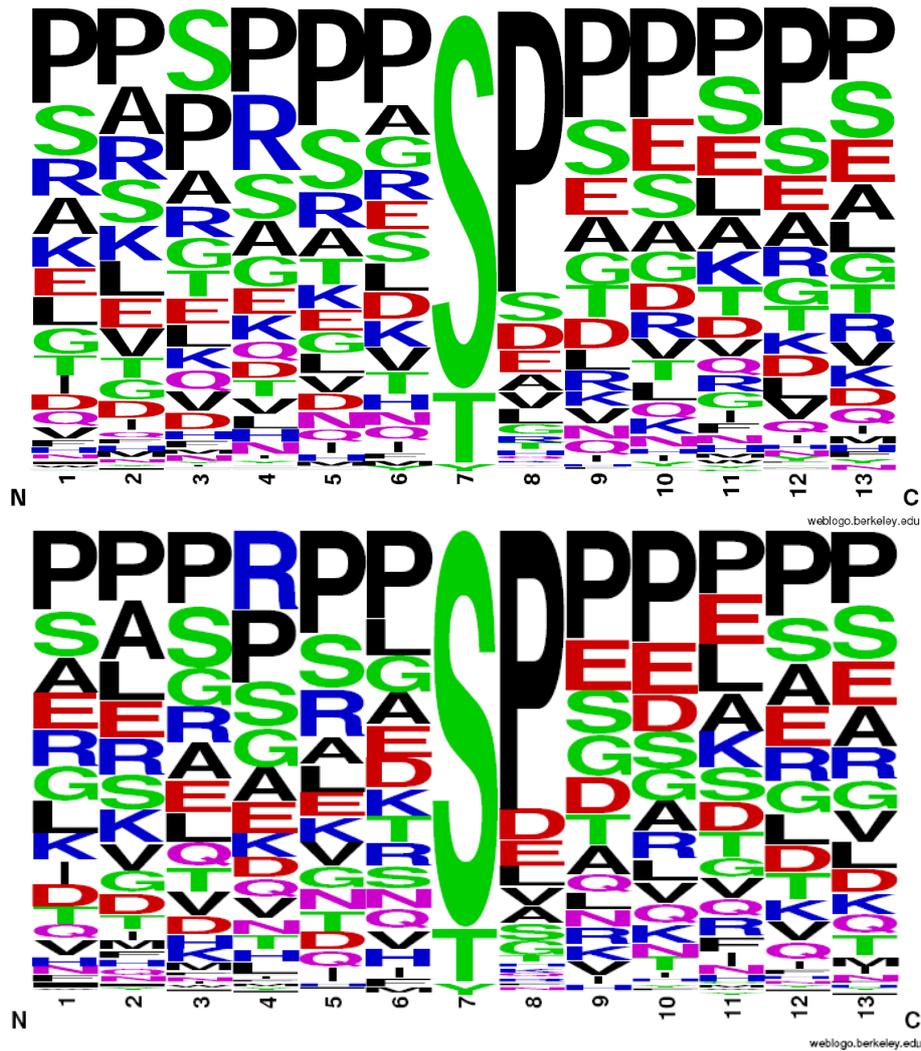


Figure 4.2.10 WEBLOGO (<http://weblogo.berkeley.edu/logo.cgi>) evaluates kinase substrate motif for the down-regulated phosphorylation sites by both 5 nM and 50 nM dasatinib. The frequency plot is depicted here. Upper panel: All 424 down-regulated sites were assessed. Lower panel: within the 424 phosphosites, those identified from singly phosphorylated peptides were assessed.

We submitted all the sequence windows for 424 phosphosites at once (Figure 4.2.10 upper panel), as well as selectively submitted those down-regulated sites identified from singly phosphorylated peptides (Figure 4.2.10 lower panel). The resulting frequency plots are very similar. Most strikingly, proline-directed phosphorylation (pS/pT-P) accounts for more than half of the phosphorylation events. The web logos suggest the substrate motifs of ERK1/2 and AKT. To confirm this, we manually filtered the sequence window for the classical ERK1/2 motif (V/PXpS/pTP) and ATK motif (K/RXRXXpS/pT). Indeed, 91 sequences passed the selection for ERK1/2 substrates and 26 for ATK substrates. These numbers indicate ERK1/2 and AKT signaling cascades make up a significant portion of BCR-ABL signaling pathway.

4.2.2.3 Effect of dasatinib on the overall BCR-ABL signaling pathway

The accumulating knowledge in BCR-ABL signaling has revealed several critical pathways that contribute to chronic myeloid leukemia transformation²²⁶. In association with SHC and GRB2, BCR-ABL activates ERK1/2 and JAK/STAT pathways and therefore causes cells to bypass the growth factor-dependency of proliferation and cell growth. BCR-ABL activates the PI3K-AKT and JAK/STAT pathways to enhance cell survival, while its activation of focal adhesion components (actin, FAK, CAS, etc) leads to a decrease in cell adhesion and abnormal interaction with extra-cellular matrix and stroma.

According to the summary from Weisberg *et al*²²⁶, we recapitulated the BCR-ABL signaling network with details of detected, down-regulated phosphosites (Figure 4.2.11). These signaling molecules distribute among the various aforementioned pathways.

Critical proteins involved in the three major MAPK cascades were suppressed, including the autophosphorylation loop of p38 α and ERK1/2. This is in agreement with the ERK1/2 kinase substrate motif analysis, where a large number of suppressed phosphosites display the classical substrate motif for ERK1/2.

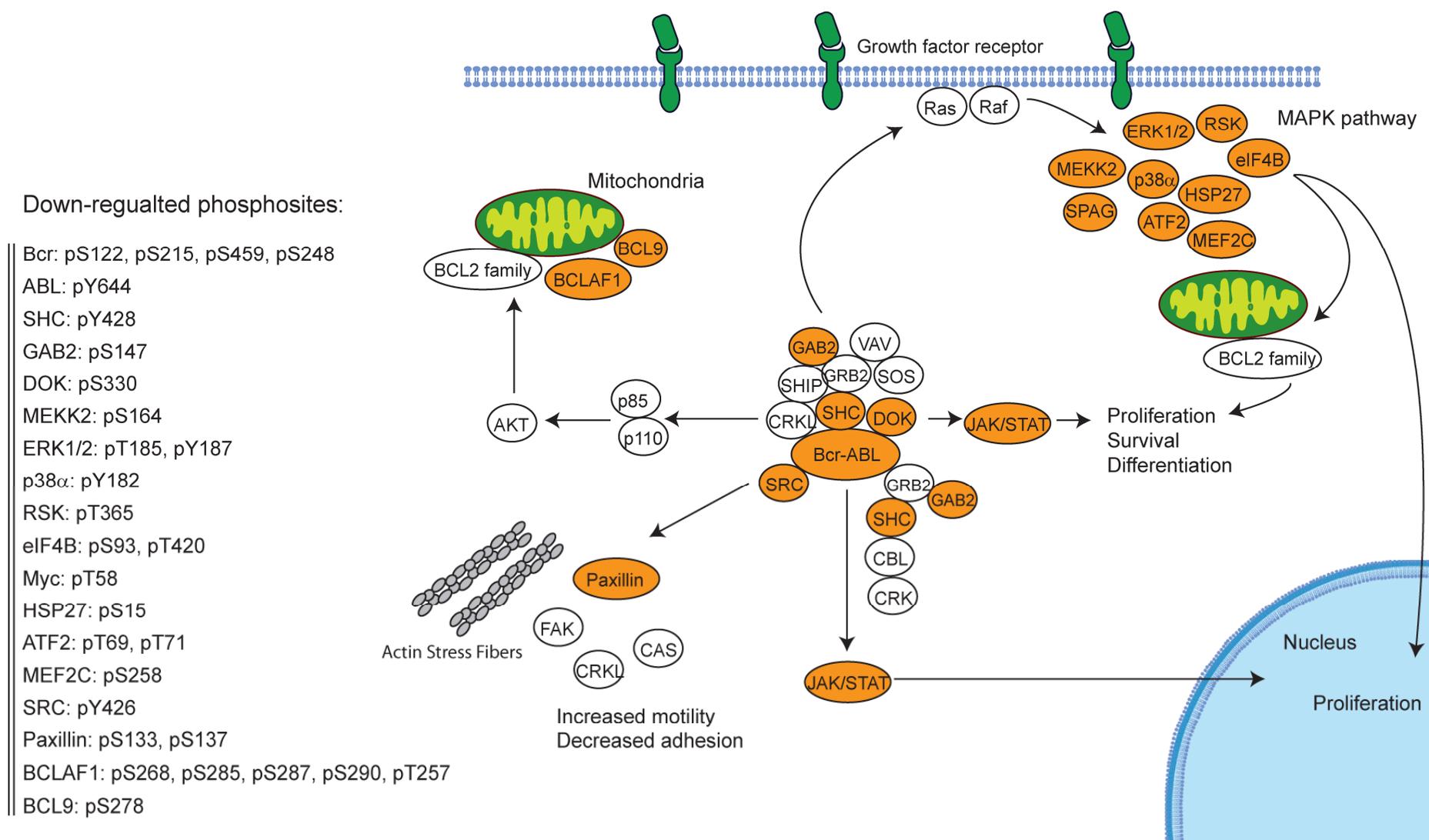


Figure 4.2.11 Effect of dasatinib on the BCR-ABL signaling pathway. The phosphorylation sites that are suppressed by both 5 nM and 50 nM dasatinib are displayed on the left. Signaling pathways are based on Weisberg *et al.*²²⁶.

Previous studies provide increasing evidence that a major mechanism of BCR-ABL transformation is via suppression of apoptosis²²⁷. Primary CML progenitors required growth factors for proliferation²²⁸. Yet in the absence of serum or growth factors, primary CML progenitors had increased cell viability compared to normal cells²²⁹. Antisense oligonucleotide-mediated down-regulation of BCR-ABL conferred susceptibility of BCR-ABL positive cells to apoptotic stimuli²³⁰ and did not affect cell cycle progression²²⁹. We captured several sites of the BCL superfamily proteins as well as the p53 binding protein 2, an essential regulator of p53 in apoptosis and cell growth.

Overall, phosphopeptides from 33 phosphatases and 130 kinases were quantified. By applying the criterion that at least one phosphorylation has to be suppressed by half, 6 phosphatases and 27 kinases are selected. Dasatinib is known to bind to a large number of kinases. Carter *et al* tested a panel of 148 kinases and dasatinib influenced 76 of them²³¹. In our dataset – even though the selection was broaden to 25% suppression (Ratio \leq 0.75) – only 14 phosphatases and 51 kinases were influenced by dasatinib. This may generally reflect the difference between cell line studies and *ex vivo* enzyme assays. If our observation with dasatinib proves to be more widely applicable, it suggests that off target effects from in-vitro effects may not be reflected in cells in all cases. One obvious reason for this could be that off-target effects depend on the concentration of kinases and substrates. These are usually normalized in-vitro, whereas our assay uses the ‘correct’ cellular concentrations. Therefore, if there is an off-target effect on a very minor cellular constituent, this may not have an appreciable effect on the overall signaling network. Additionally, different cell types retain their cell-type specific features to different degrees and therefore may not employ all possible in-vivo substrates to respond to a stimulus. Some of the in-vitro targets may not even be expressed in the cell type under study.

These observations indicate that the two major themes, i.e. to study direct binding targets of the inhibitor compounds and to investigate their influence on cellular signaling network, should be combined. Together we will provide a much more complete picture of the drug effects and mechanism of actions.

4.2.3 Conclusions and discussions

In this project we quantified the effects of three kinase inhibitors, including a clinical drug, on the phosphorylation changes in the entire cellular signaling network. Two of the inhibitors, U0126 and SB202190, are very commonly employed in bench research to elucidate the cellular functions of their target kinases. To our knowledge there has been little investigation on the overall effects of these inhibitors on the general signaling network and some of the derived functions may actually be induced by secondary signaling molecules rather than the targeted kinases. Therefore interpretation of the cellular assays must be done with caution.

We derived a simple EGF signal propagation model based on the suppression of phosphorylation by selective MAPK inhibitors, which indicates that MAPK signaling cascades are dominant in the early phase of EGF signal transduction, i.e. after 15 minutes of 150 ng/ml EGF treatment. Bearing in mind that U0126 and SB202190 may not only suppress their designed targets (MEK1/2 and p38 α/β , respectively), this model is only a rough estimate. We applied the same strategy of SILAC-based quantitative phosphoproteomics to inspect the cellular effects of dasatinib, a clinical drug for CML therapy. Two different concentrations were employed to enhance the reliability of the substrate list. The influenced signaling molecules distribute among several critical signaling pathways for BCR-ABL induced CML transformation. Specific sites from these signaling molecules were pinpointed, providing site-specific information of the inhibitory mechanism.

Phosphorylation is a critical modification for various signal transduction events in normal and transformed cells. Selectively inhibiting the crucial kinases or phosphatases has proven a useful strategy for correcting the abnormal enzyme activities which can distort cellular signaling networks. Establishment of this novel type of drugs in the clinic is represented by imatinib, dasatinib, erotinib, and others. As the majority of studies so far have focused on the direct binding targets of these inhibitor compounds, there is a need to elucidate their global influence on signal transduction. Our quantitative phosphoproteomics approach is able to examine the cellular effect at the phosphorylation and - in principle - at the protein expression level. Thus our datasets complement the mechanism of actions previously described for the three investigated kinase inhibitors.

Since the first kinase-targeted inhibitor compound was successfully developed and applied to patients, clinicians have been constantly in battle with disease evolution. Besides elevated protein expression levels, point mutations have been frequently reported in the targeted

kinases. This drug-resistance has motivated scientists to screen more compound libraries and further develop inhibitor compounds. To meet the need of testing the efficacy and dynamics of the new compounds in an unbiased way and on a global scale, we believe our simple but powerful SILAC-based quantitative phosphoproteomics technique can greatly facilitate the screening processes from the early phase of the drug development.

Two interesting concepts have started to emerge in the drug development field: that multi-kinase inhibitory compounds can be effective drugs and that the combination of multiple inhibitors in one therapy can lead to much better clinical results. For example, sunitinib exerts antitumor activity and induces tumor regression in cancer patients via inhibition of growth promoting signaling in tumor cells (e.g. c-Kit, Flt-3, RET) and of receptors expressed on cells of stroma (e.g. VEGFR, PDGFR)^{75, 232}. Combination of pharmacological inhibitors of PI3-K (LY294002 and Wortmannin) with imatinib induced more apoptosis in both chronic and blast crisis CML cells²³³. The T315I point mutation in BCR-ABL conferred resistance to imatinib and dasatinib. However, a PDK-1 inhibitor (OSU-03012), which essentially inhibits Akt activation, can synergize with imatinib in inducing apoptosis even in cells expressing the T315I BCR-ABL. As more and more complex inhibitory therapies are attempted, the cellular effects of these inhibitor compounds will become even more complicated. With our unbiased, high-throughput, and robust technique, we believe that knowledge of principles of drug actions can be derived relatively fast and that this can guide future drug development.

4.3 Project 3 - 'Proteomic phenotyping' to assess differences between transformed and non-transformed mouse liver cells

This work is included in a submitted manuscript:

Comparative proteomic phenotyping of cell lines and primary cells to assess preservation of cell type specific functions

Pan, C¹; Kumar, C¹; Bohl, S; Klingmueller, U; and Mann M.

¹ both authors contributed equally

The cell line Hepa1-6 was isolated from the mouse liver with hepatoma¹⁷⁰, therefore it is thought to retain the characteristics of both mouse hepatocytes and cancer (Section 3.1.1). In this project we compared the proteome of freshly isolated hepatocytes from healthy mice and the proteome of Hepa1-6, with the aim of characterizing their differences at a systematic and quantitative level. Striking differences in protein expression, molecular function and signal transduction pathways were revealed. Clearly, cancer-derived cell lines are mostly dedicated to proliferation and growth, representing major characteristics of cancer. The primary hepatocytes, on the other hand, maintain their functions in metabolism and organ context. Proteome quantitation in this study helped to reveal these functional differences at the level of the proteome for the first time. This should help to better evaluate research results derived from this cell line. More generally, this study alerts the biologists to where cell lines may significantly distort intracellular network and therefore where caution should be applied when analyzing cell line results. Fortunately, the proteomic phenotyping strategy (Figure 4.3.1) presented in this work is very robust. It can be applied to studies like the one presented here, i.e. comparing a cell line and their cognate primary cells, but is not limited to this type of study. Due to the strength of the novel bioinformatic algorithm (section 3.5.1 and section 4.3.2) developed in this work, even proteomes originating from distinct sources can be compared quantitatively and systematically in terms of specific cellular functions.

4.3.1 Proteomes of primary hepatocytes and Hepa1-6

To characterize phenotypic differences between cell lines and primary cells, we labeled the murine hepatoma cell line Hepa1-6 with Arg10 ($\text{Arg}^{13}\text{C}_6^{15}\text{N}_4$) and Lys8 ($\text{Lys}^{13}\text{C}_6^{15}\text{N}_2$), i.e. heavy forms of the amino acids used in SILAC labeling. Primary hepatocytes were isolated from six to twelve weeks old B6 mice according to standard operating procedures established by the German systems biology competence network HepatoSys. After isolation primary hepatocytes were cultured in dishes for 14 h. Proteins from completely labeled Hepa1-6 and primary hepatocytes were extracted and mixed in equal amounts (section 3.1.5). 100 μg of proteins from each of two mixtures were separated by isoelectric focusing and analyzed in the LTQ-FT using the top 5 LC-MS/MS method (section 3.2). For biological and technical replicates, we obtained primary hepatocytes from two mice and treated the samples separately during the experiments. Data analysis was done according to section 3.4.1.

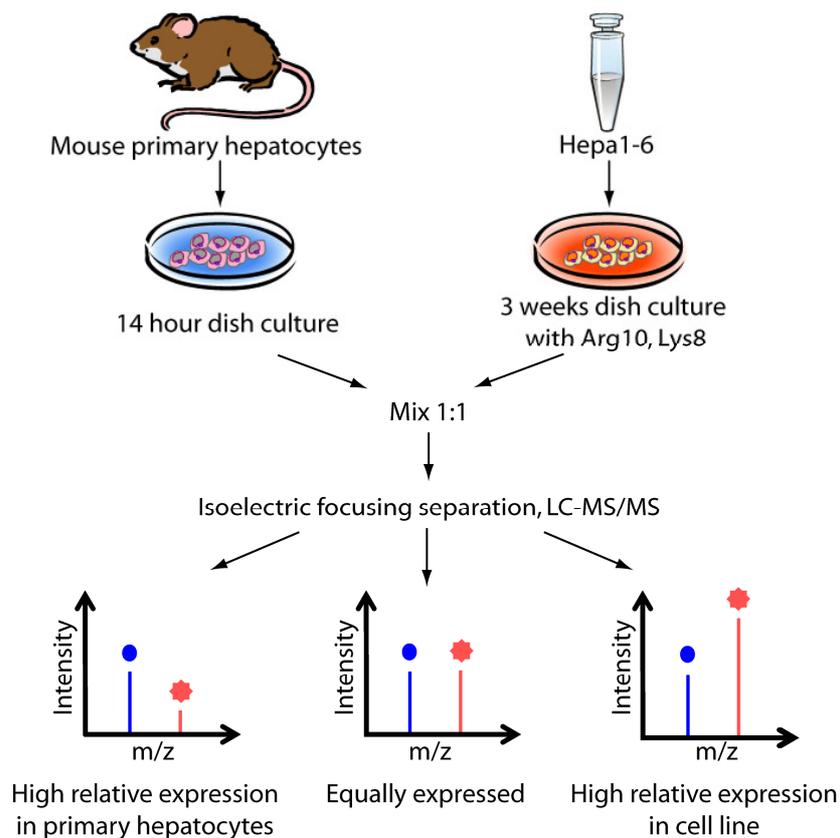


Figure 4. 3.1 **Strategy for comparing primary cells with immortalized cell lines.** SILAC-labeled Hepa1-6 cell line was combined with mouse primary hepatocytes. After cell lysis and digestion peptide mixtures were separated by isoelectric focusing (Offgel) and analyzed by LC/MS-MS using an LTQ-FT mass spectrometer.

4.3.1.1 Excellent reproducibility of the identification and quantitation of the proteomes

We used high resolution mass spectrometry to identify and quantify around 4,000 proteins in each mixture of primary hepatocytes and Hepa1-6. The overlap of identified proteins is very large between the two datasets (Figure 4.3.2, inset). As quantitation in each dataset was performed, we also wanted to evaluate if the relative protein expressions were reproducible between the biological replicates. With 100% similarity set to be 1, the Pearson correlation coefficient of these two datasets is 0.95 (Figure 4.3.2). Given that these two primary hepatocyte samples were obtained from different mice, this correlation coefficient is excellent. On one hand, this reflects the robustness and stability of our experimental protocols and particularly the LC-MS/MS system. On the other hand, the homogenous genetic background of these B6 mice also contributed greatly to the reproducibility.

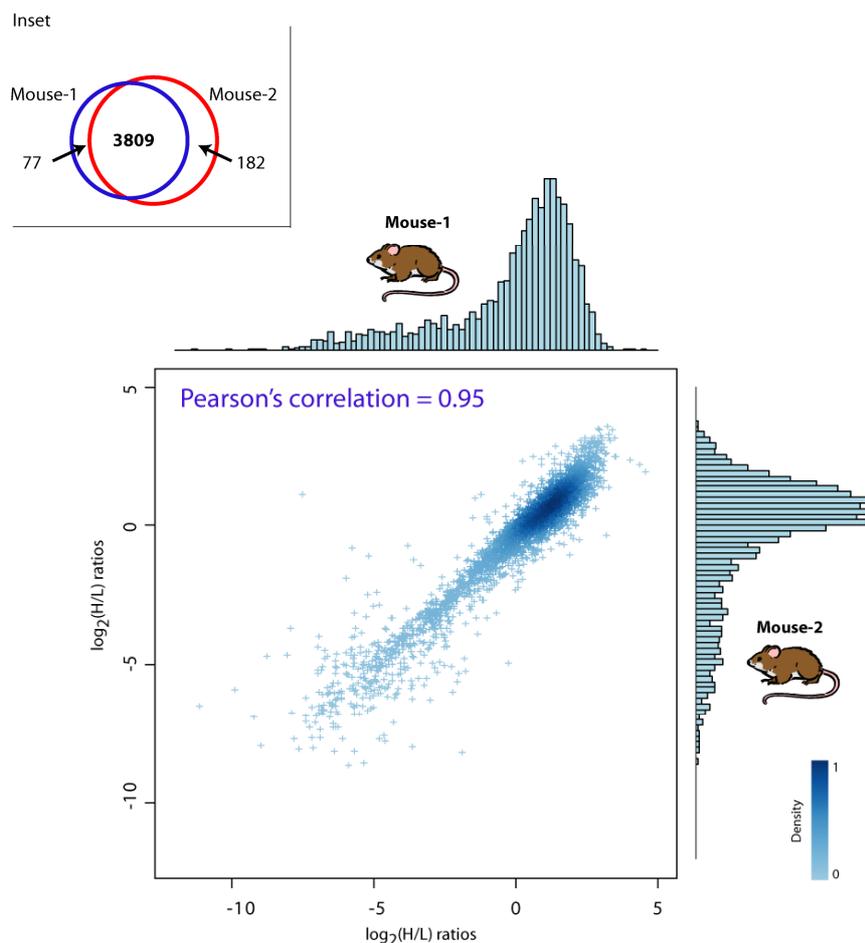


Figure 4.3.2 Replicate experiments of comparing Hepa1-6 cell line with primary hepatocytes from two mice achieved very high degree of reproducibility at both identification level (inserted in the upper-left corner) and quantitation level (about 4,000 proteins quantified; Person correlation coefficient 0.95).

We then combined the two datasets and analyzed them together in MaxQuant using stringent and unified criteria (section 3.4.1). At a false positive rate of less than one percent, a total of 4,089 proteins were identified and quantified between the two cell populations. In the following, unless otherwise specified, the analyses were performed on the two proteome datasets separately.

4.3.1.2 Drastic changes in protein expressions in primary hepatocytes and Hepa1-6

We compared the expressions of proteins that were present in both primary hepatocytes and Hepa1-6. The primary and cell line proteomes overlap qualitatively but are very different quantitatively, with more than half of the proteome changing at least two-fold between the two conditions (Figure 4.3.3).

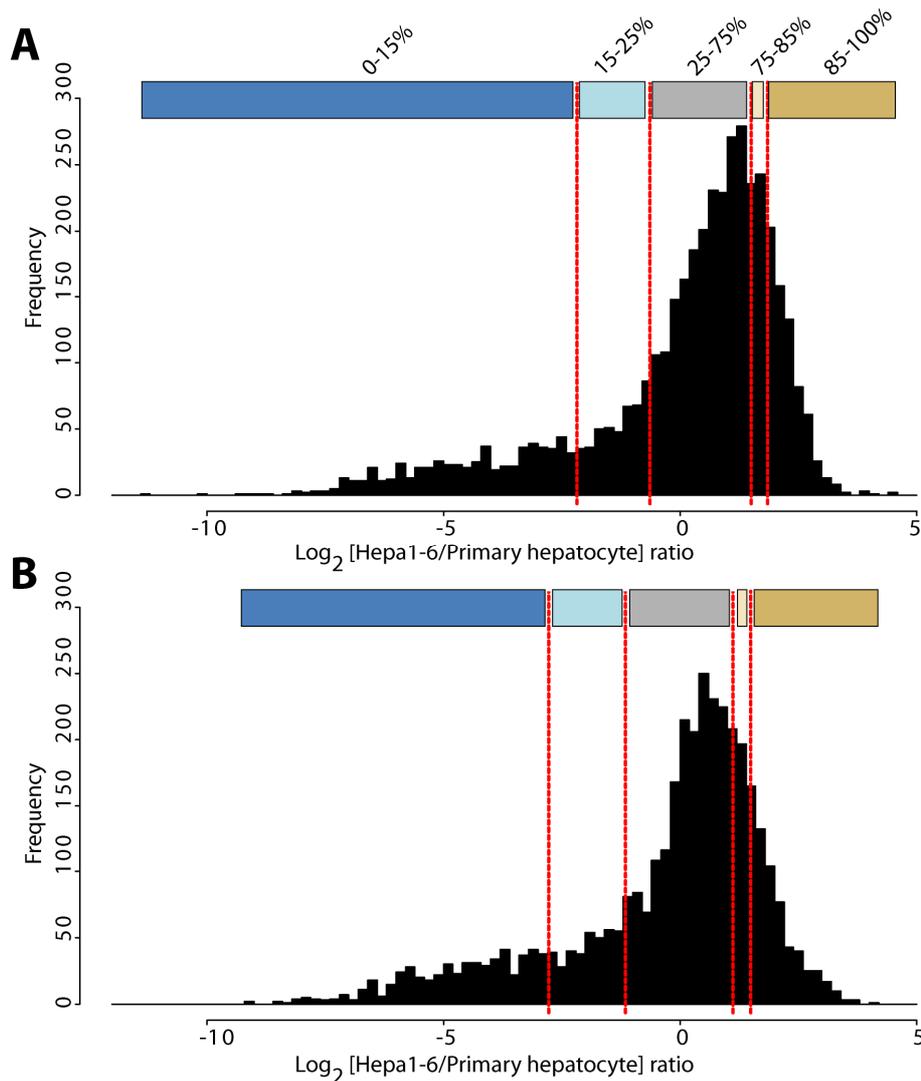


Figure 4.3.3 Fold-change distributions of the proteome. **(A)** Quantitative comparison of the Hepa1-6 cell line vs. the primary cell proteome. The distribution was divided into five quantiles as follows. High relative expression in primary cells (0-15%, at least four-fold down-regulation), mostly expressed in primary cells (15-25%, -4 to -1.5 fold regulation), not highly regulated proteins (25-75%; -1.5 to +2.8), mostly expressed in Hepa1-6 (75-85; 2.8 to 3.6 fold), highly expressed in Hepa1-6 (85-100%, more than 3.6 fold change). Color coding of these categories is indicated at the top of the panel. **(B)** Biological replicate of the experiment showing excellent reproducibility (see figure 4.3.2).

Many proteins are expressed at much lower levels in the immortalized cell line than in the primary cells whereas comparatively few were up-regulated in Hepa1-6. This is surprising since cancer cells are thought to be de-differentiated and to express many genes inappropriately.

4.3.1.3 Changing glucose level in cell culture does not induce drastic proteome changes

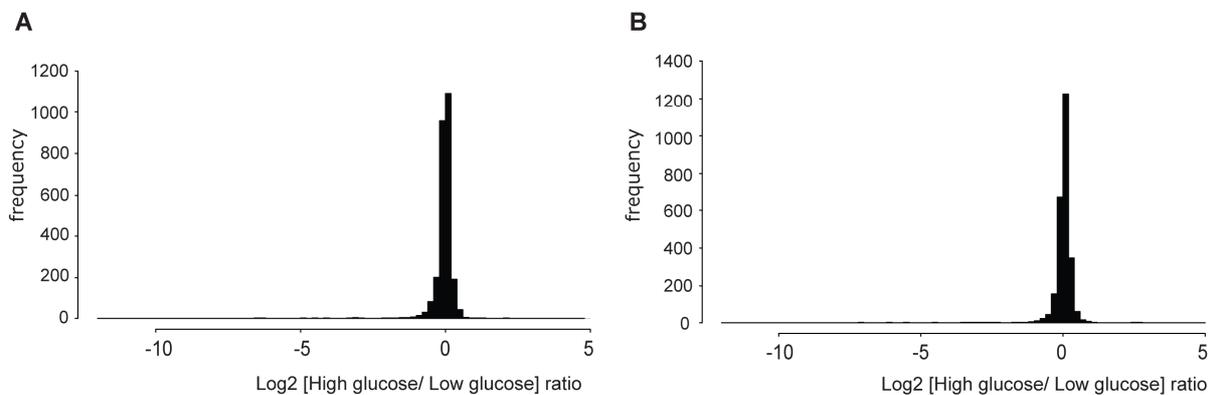


Figure 4.3.4 Quantitative comparison of the Hepa1-6 proteome cultured in high glucose (38 mM) and physiological glucose concentration (8 mM). Two independent comparisons **(A)** and **(B)** were performed starting from cell culture and SILAC labeling.

It is usually recommended to cultivate Hepa1-6 cells in high glucose medium (38 mM). Therefore we asked whether some of the observed phenotypic changes are attributable to this circumstance. To address this experimentally, we performed another SILAC experiment

comparing Hepa1-6 cells cultured in high glucose against cells cultured in physiological glucose levels in mice (8 mM) for three weeks. In this experiment, there were hardly any overall changes in the proteome and 96% of the proteins were of constant abundance within a factor of 1.5 (Figure 4.3.4 A). This was also confirmed in a replicate experiment (Figure 4.3.4 B). These results rule out a dominant role of the super-physiological glucose level in the proteome differences between primary cells and cell lines. Furthermore, they demonstrate excellent quantitative accuracy of our experiment on a proteome-wide basis.

4.3.2 Phenotyping of the proteomes of primary hepatocytes and Hepa1-6

4.3.2.1 Functional clustering on top of quantitative clustering - a method for proteomic phenotyping

To functionally understand the differences between the two cell populations, we divided the fold-change distribution between primary hepatocytes and the Hepa1-6 cell line into five quantiles according to relative protein expression (Figure 4.3.3). Each quantile was assessed separately for overrepresented pathways, biological processes and cellular components with Gene Ontology (GO) and KEGG pathway analysis^{234, 235} (Section 3.5.1). We retained each functional category that reached at least 95% statistical significance in one of the quantiles and then performed one-way unsupervised clustering of the p-values of the resulting categories. This analysis differs from the more familiar clustering of overrepresented genes themselves, which is frequently employed in microarray-based experiments. It integrates the strength of statistical testing (taking p-values as input for clustering) with the intuitive simplicity of hierarchical clustering. By automatically classifying related processes and pathways based on their up or down-regulated protein measurements, it provides an unbiased global portrait of representative biological functions, enabling visual interpretation of the phenotype in terms of aggregate functional modules on a systems level.

4.3.2.2 P-value clustering on the proteomes of primary hepatocytes and Hepa1-6

We applied this novel p-value clustering strategy to our two proteome datasets. The result of KEGG pathway over-representation analysis for one of the dataset is shown in figure 4.3.5 and Appendix 4. For the same dataset, results from the cellular component category in Gene Ontology, and the biological function category in Gene Ontology are displayed in Appendix 5

and 6. We verified the robustness of these functional assignments by comparing the shared p-value matrix of the replicate experiments against each other (Appendix 7-9). This correlation was 0.86 for KEGG, 0.85 for GO biological process and 0.92 for GO cellular compartment.

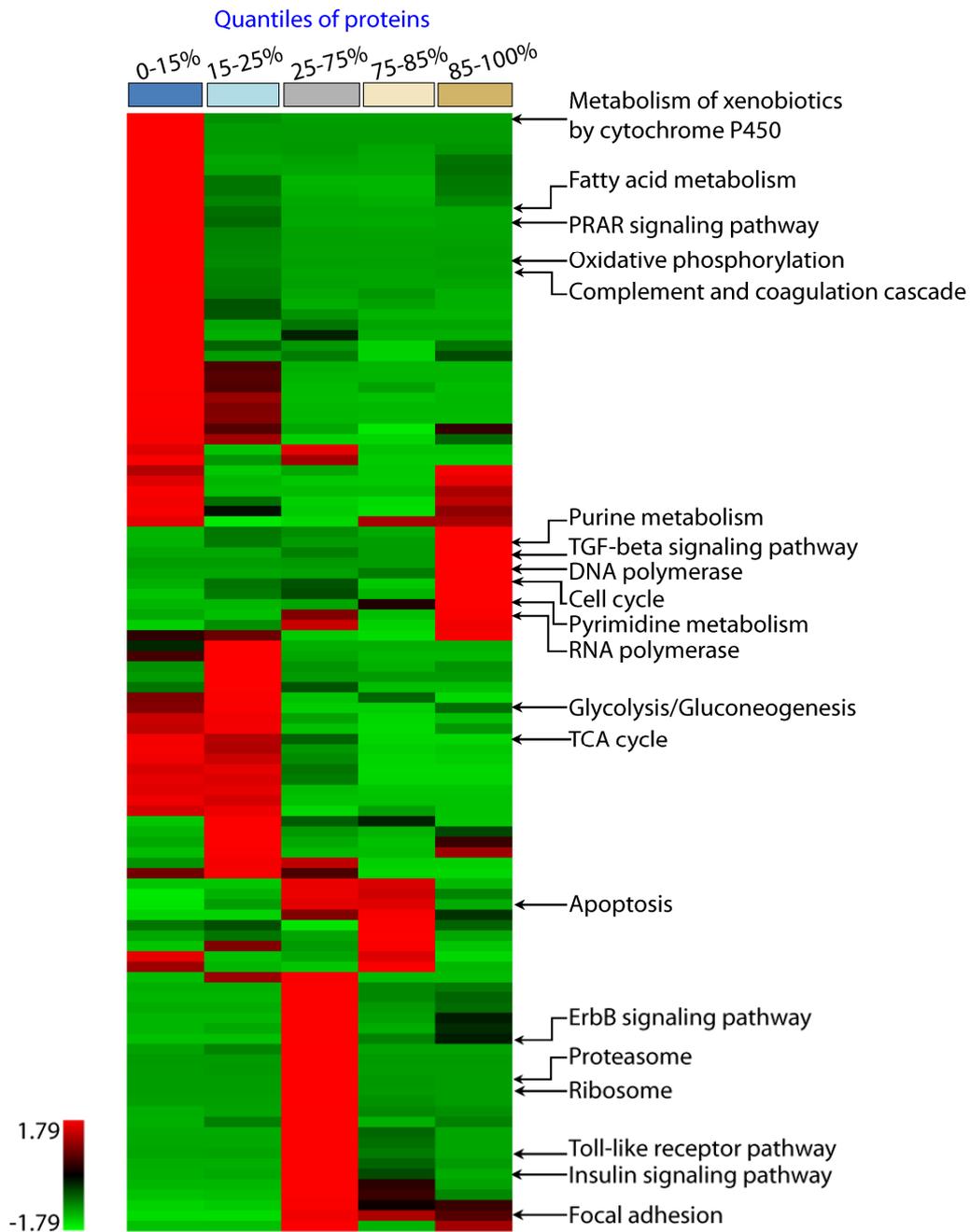


Figure 4.3.5 Functional phenotyping of the proteome. The five quantiles (see Figure 4.3.3) were separately analyzed for enriched KEGG pathways and clustered for the z-transformed p-values. The color bar on top represents the quantiles. Representative pathways enriched in the protein population of each quantile are annotated. For complete listing of significant categories and p-values see Appendix 4.

4.3.2.3 Cell growth related proteins are highly expressed in Hepa1-6

The most prominent cluster of proteins expressed at higher levels in Hepa1-6 relates to cell division and encompasses categories such as cell cycle ($p < 10^{-9}$), DNA synthesis ($p < 10^{-4}$) and RNA polymerase ($p < 10^{-3}$). This cluster consists of 10 enriched pathways, of which at least five relate to increased cell proliferation. Relative expression levels of proteins involved in the cell cycle are depicted in Figure 4.3.6. Biologically, this is not surprising since hepatocytes in the liver and in our primary culture are largely arrested in the G₀ phase of the cell cycle, whereas Hepa1-6 cells double every 18 hours. Nevertheless, the fact that this phenotypic trait is so clearly grouped in the cluster analysis makes it an excellent positive control.

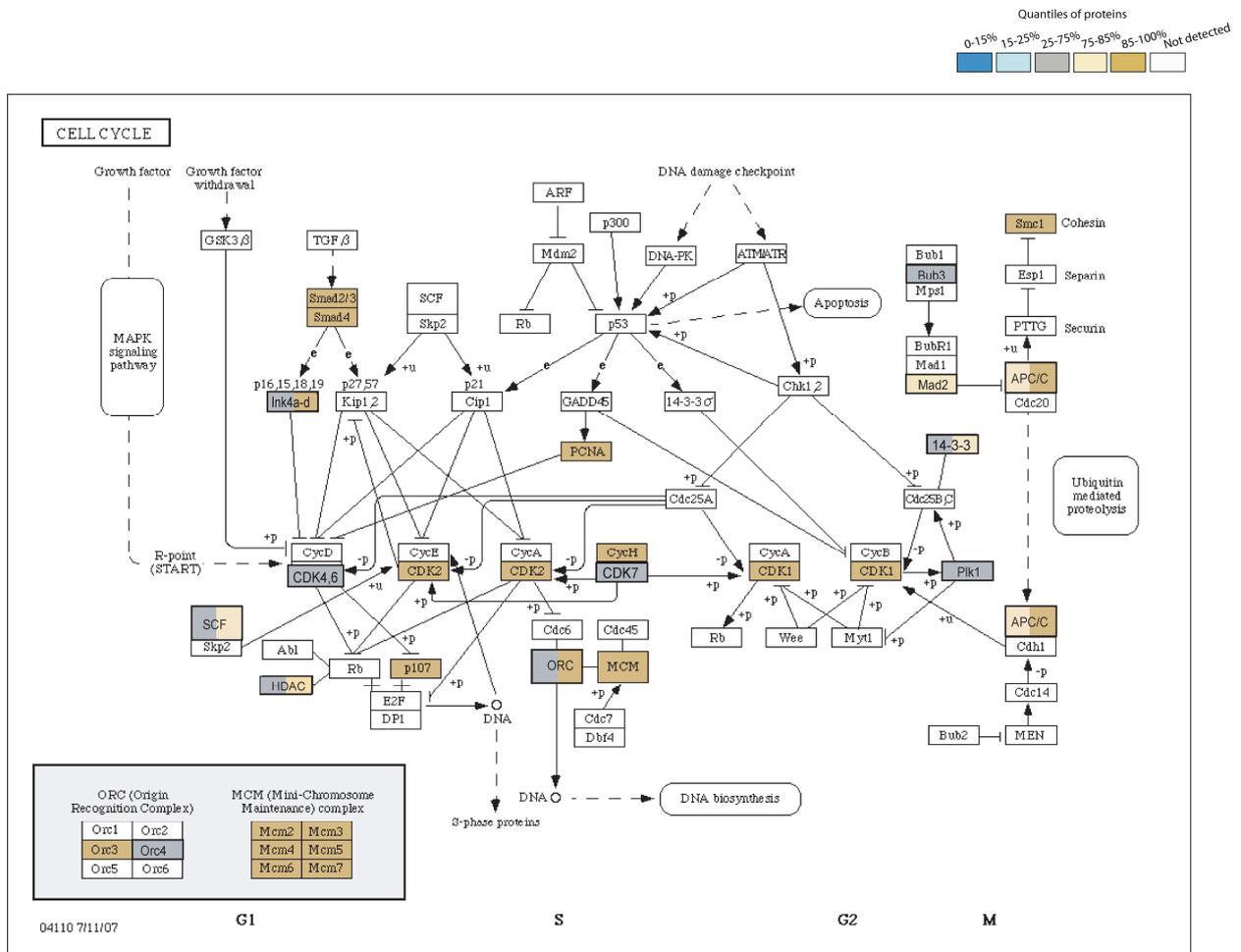


Figure 4.3.6 KEGG pathway mapping of cell cycle. The color bar on top represents the quantiles. Proteins that are marked with two colors are likely to represent isoforms.

TGFβ-mediated signaling is more highly represented in the Hepal-6 cell line and the canonical members TGFβ R1, Smad2/3, Smad4, p107 and p15 are all up-regulated significantly (Figure 4.3.7). This was unexpected because TGFβ is usually associated with growth inhibition whereas Hepal-6 has increased proliferation rate compared to primary hepatocytes. However, the biological actions of TGFβ are complex and it is thought to shift from a growth inhibitory to a growth promoting role during cancer development²³⁶. Thus up-regulation of this pathway suggests that in the Hepal-6 tumor cells, TGFβ may have growth promoting effects. Taken together, our data indicate that biological functions related to many important signaling pathways are well preserved in Hepal-6.

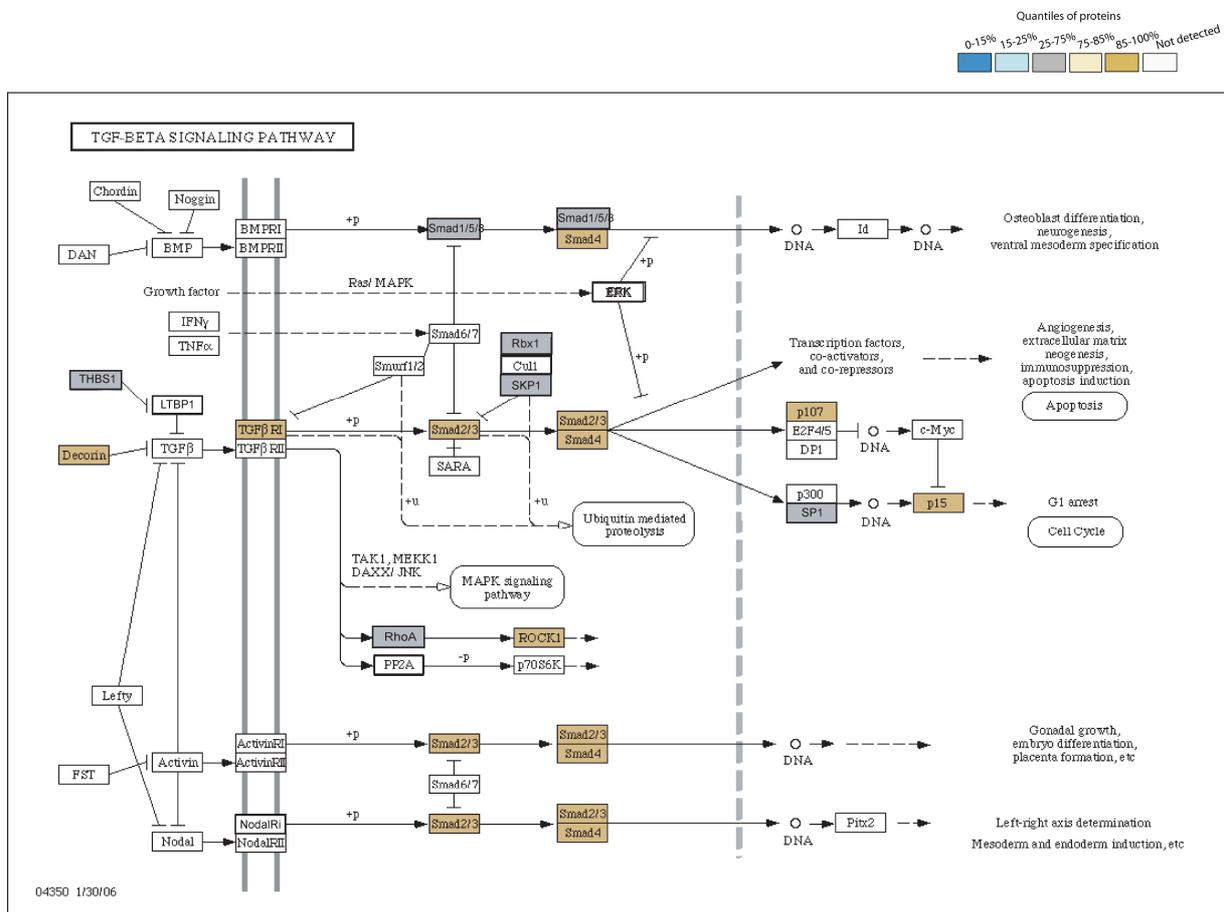


Figure 4.3.7 Proteomics phenotyping at the pathway level. KEGG pathway mapping shows that TGFβ signaling pathway is predominantly present in the cell line.

4.3.2.4 Proteins relevant to liver functions are highly expressed in primary hepatocytes

Drug metabolizing enzyme families are highly expressed only in primary cells.

One of the most enriched categories in the quantile most expressed in primary cells is the P450 family of enzymes ($p < 10^{-16}$). These enzymes are mainly involved in metabolizing endogenous substances and xenobiotics²³⁷, a prototypical function of the liver. We identified 32 different P450 proteins and 25 of them were down-regulated at least tenfold in the cell line. Furthermore, the flavin monooxygenase (FMO), UDG-glucuronosyltransferase (UGT), sulfotransferase (SULT), and glutathione S-transferase (GST) - additional prominent drug metabolizing enzyme families (DMEs) - were also severely down-regulated in Hepa1-6 (Appendix 10). Only three P450s were up-regulated. Two of them (CYP1A1 and CYP2S1) are known to be regulated by the aryl hydrogen receptor^{238, 239}. This receptor was also more highly expressed in Hepa1-6, providing a ready explanation for the up-regulation. The third up-regulated P450 protein (RIKEN clone E130013F06) has only been characterized on the basis of sequence homology and may have functions different from traditional P450 enzymes.

Reduction of DME activity is a notorious difficulty in toxicological assays in cell lines. Toxicologists therefore attempt to stimulate liver cell lines with the aim of boosting DMEs activity²⁴⁰. Quantitative knowledge of the changes in the profile of DMEs could provide a rational basis to adapt cell systems to more closely mimic hepatocytes *in vivo*.

Proteins related to liver organ context were highly expressed.

Another prominent and cell-specific function of hepatocytes is production of plasma proteins. Figure 4.3.5 reveals that 'complement and coagulation cascade' is specific for the primary cells ($p < 10^{-2}$). Inspection of the pathway involved (Appendix 11) shows that major liver-produced factors, such as C3, C4, MBP-C, F2, F5, A2M, Serpin A1/C1 and apolipoproteins are down-regulated more than five-fold in Hepa1-6 compared to primary hepatocytes. Interestingly, p-value clustering in cellular component category revealed a heavy overrepresentation of extracellular matrix ($p < 10^{-18}$) in primary hepatocytes (Appendix 5 and 8). Apparently, the cell line also shows lower expression of proteins related to communication with stroma and with tissue maintenance. Thus, it is likely that loss of tissue context allows the cell line to shut down this function, which is nonessential for propagation in culture.

Mitochondrial and other metabolic related proteins are highly expressed in primary hepatocytes.

In the cellular component p-value clustering, mitochondria are indicated as the most overrepresented group in the primary hepatocytes ($p < 10^{-62}$).

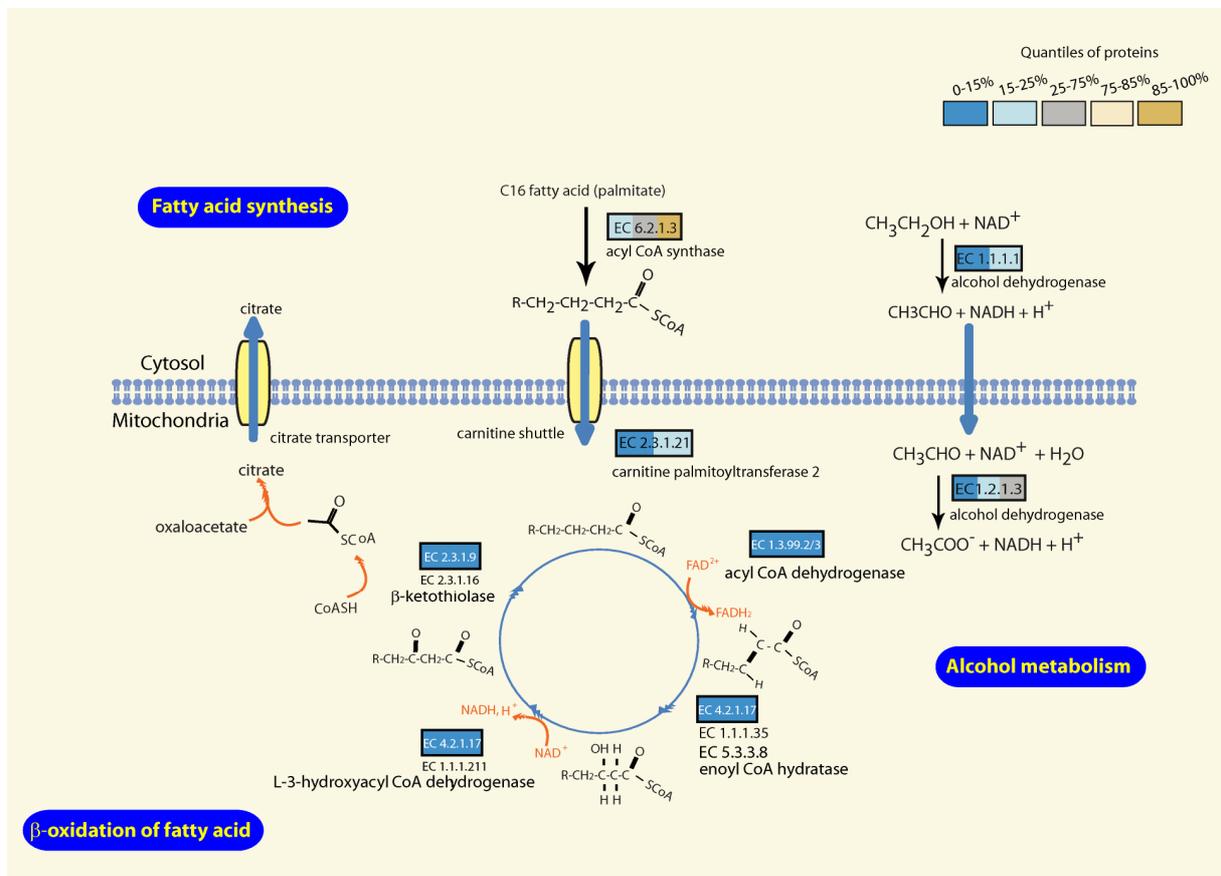


Figure 4.3.8 Phenotypic proteome comparison at the pathway level. Mapping of protein ratios on the fatty-acid metabolism pathway reveals that almost the entire module is down-regulated several-fold in Hepal-6. Proteins are color-coded according to their relative expression in the two cell types according to the scheme in Figure 4.3.3. Proteins that are marked with two colors are likely to represent isoforms.

It is known that hepatocytes contain 1000–2000 mitochondria per cell making up one fifth of the cell volume. Consistent with this feature, we identified a large number of mitochondrial proteins. The IPI mouse database used in this study contains 990 proteins with mitochondrial annotations. A total of 479 mitochondrial proteins study were found in our proteome, comprising around 12% of the proteome. Of these identified mitochondrial proteins, 69% were in the asymmetric tail of the distribution, indicating they were expressed several fold lower in Hepa1-6 cells than in primary hepatocytes. These proteins carry out well known mitochondrial functions, such as fatty acid synthesis, β -oxidation, and alcohol metabolism (Figure 4.3.8).

We independently confirmed this observation by DAPI and Mitotracker staining (Figure 4.3.9).

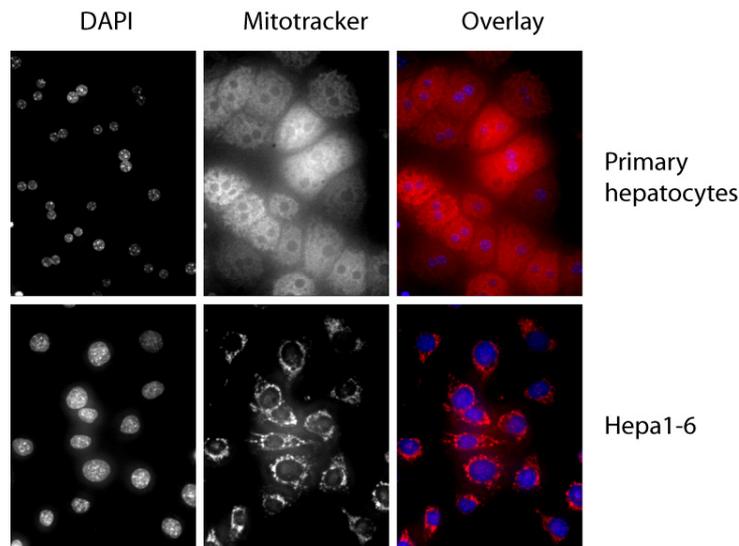


Figure 4.3.9 Nuclear (DAPI) and mitochondrial (Mitotracker) staining of primary hepatocytes and Hepa1-6 cells. Many primary hepatocytes are binuclear²⁴¹.

Indeed, primary hepatocyte nuclei were much smaller whereas in these cells mitochondria were much more abundant with respect to Hepa1-6. Concurrent with this, fatty acid metabolism was drastically down-regulated in Hepa1-6 according to enrichment analysis of KEGG pathways (Figure 4.3.5). Likewise, ‘oxidative phosphorylation’ ($p < 10^{-29}$), ‘urea cycle’ ($p < 10^{-4}$) and ‘steroid biosynthesis’ ($p < 10^{-2}$) were statistically significantly enriched in the quantile most expressed in primary hepatocytes. These down-regulated metabolic functions at least partially take place in mitochondria. Conversely, parts of the glycolysis pathway were up-regulated in Hepa1-6 (Appendix 12). Together, our results portray a drastic

metabolic rearrangement, away from oxidative metabolism in the mitochondria and towards less efficient anaerobic metabolism. As discussed further below, these findings provide evidence for the Warburg hypothesis, that cancer cells shift towards glycolytic metabolic pathways²⁴².

4.3.2.5 Proteins involved in several signaling pathways were expressed to similar degrees in primary hepatocytes and Hepa1-6

In the category containing the 50% of proteins with the least change, many household functions and organelles including ribosome ($p < 10^{-2}$), proteasome ($p < 10^{-3}$), splicing ($p < 10^{-4}$) and Golgi apparatus ($p < 10^{-3}$) are significantly enriched. Interestingly, several signaling pathways are also preferentially located in this quantile. These include the ErbB and PI3K signaling pathways (Figure 4.3.10). This finding is in agreement with the requirement of growth factor containing serum for the maintenance of most cell lines.

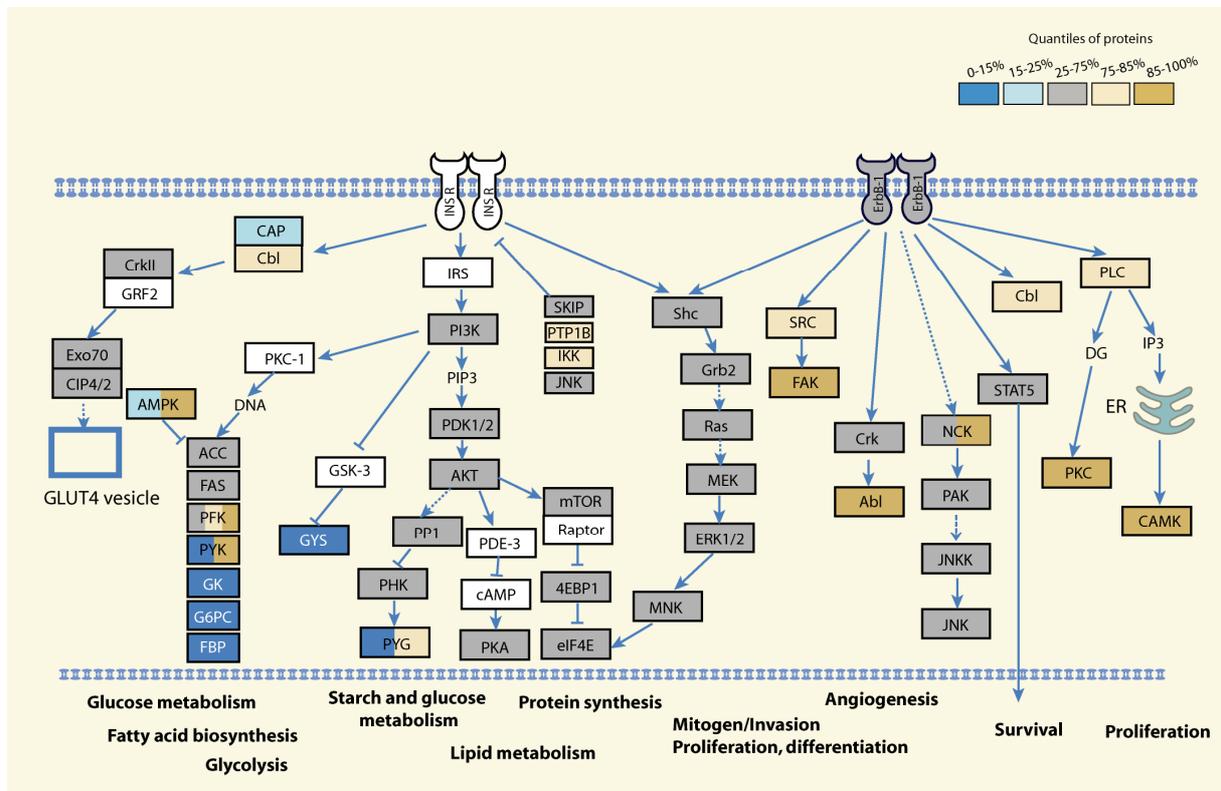


Figure 4.3.10 Phenotypic proteome comparison at the pathway level. KEGG pathway mapping of ErbB and PI3K signaling pathway shows that they are equally present in primary cells and the cell line.

4.3.2.6 Other observations

Some categories shared by both cell types and enriched when analyzed using the KEGG database represent non-liver functions (such as ‘long term potentiation’) or even non-animal functions (such as ‘CO₂ fixation’). However, the enzymes found in these categories function both in liver tissue as well as in neurons or plants. For example, all identified CO₂ fixation proteins also function in citric acid cycle. Most identified proteins for long-term potentiation are kinases and phosphatases involved in EGF, insulin and other signaling pathways. Therefore, overrepresentation of these ‘CO₂ fixation’, ‘long term potentiation’ and similar categories reflects the still evolving state of annotation of pathway databases rather than a limitation of our technology.

4.3.3 Conclusions and discussions

4.3.3.1 A robust and powerful technique to assess cell type preserved characteristics

Taking advantage of the ability of SILAC to compare the levels of thousands of proteins in different cellular states^{89, 243} and a novel bioinformatic approach, we have, for the first time, compared the proteomes of primary cells to cell lines. The resulting quantitative proteome of more than 4,000 proteins had an asymmetric distribution, with many proteins down-regulated in the cell line. Bioinformatic analysis of the quantitative proteomics phenotypes revealed that Hepa1-6 cells are deficient in mitochondria, reflecting re-arrangement of metabolic pathways, drastically up-regulate cell cycle associated functions and largely shut down drug metabolizing enzymes characteristic for the liver. This phenotype is ‘rational’ for rapidly dividing and not nutrient limited cells and may partly reflect Darwinian selection of cell clones.

Many biological experiments are performed in cell lines since they are readily available and readily accessible. The development of tissue culture techniques and establishment of cell lines has been ongoing for several decades. However, there are general concerns that cell lines may differ from the *in vivo* situation in important aspects. Cell lines are usually derived from tumors and have adapted to growth in culture. Thereby they may lose tissue specific functions and acquire a molecular phenotype quite different from cells *in vivo*. Thus animal experiments or studies in primary cell lines are often preferred despite their added complexity. Accurate molecular phenotypes to determine if the function to be investigated is preserved in cell lines would enable a rational choice of the most appropriate experimental system. In biotechnology

and the pharmaceutical industry this goal obtains added urgency in light of efforts to reduce animal experimentation to a minimum.

Here we introduced a straightforward technology to quantify these differences. Our proof-of-principle study in a mouse liver cell line and their cognate primary cells clearly demonstrated drastic functional differences. This quantitative knowledge of proteome changes provides an important basis to adapt cell lines to closer resemble physiological conditions.

Our technology is accurate, relatively rapid and should now allow selection of the appropriate cell system based on a global and unbiased profile according to desired biological function. Furthermore, it can be used to manipulate the cell line system to better reflect the *in vivo* situation at the proteome level. While we have based our analysis on protein expression levels, it could just as well be applied to assess fidelity of signaling pathways in cell lines using SILAC-based quantitative and global phosphoproteomics³¹.

Our analysis differs in important points from the more familiar measurement of mRNA levels by microarray and its associated bioinformatics²⁴⁴. Even though reproducibility of microarray chips has become much better during recent years, the data is not quantitative with respect to the final, desired parameter – the global change in protein levels. Furthermore, results of any specific transcript on the chip generally have to be validated by RT-PCR and then by quantitative immunoblotting. This is impractical for large numbers of proteins. In contrast, quantitative proteomics inherently contains the fold-change for each protein, and increasingly also that of specific isoforms. The quantitative nature of our results also made it possible to directly group overrepresented functions and processes instead of the evaluating the transcriptional profile.

Here we have analyzed interesting, but relatively general phenotypic, traits of two cell populations. While many of the resulting observations can be immediately rationalized in terms of biological function, they have never been quantified in a global and unbiased way. The combination of very high quantitative accuracy at the proteome level with increasingly accurate pathway databases should allow even richer assessment of the phenotypic state of any cell population in the future. Rapid advances in MS techniques and programming algorithms have made label-free quantitation more precise and accurate. Thus the algorithms described here may also be applicable to more distantly related proteomes and those that cannot be SILAC-labeled.

4.3.3.2 Molecular basis of metabolic differences between transformed and non-transformed cells

Our observation that mitochondrial proteins are drastically under-represented in transformed cells is a rediscovery of an old paradigm: the “Warburg effect”. In normal physiology glucose oxidation is composed of two steps: glycolysis in the cytoplasm, and the TCA cycle and oxidative phosphorylation in mitochondria. However in tumor cells, this principle of metabolism is violated. Back in the 1930s, the German biochemist Otto Warburg discovered that cancerous cells had elevated glycolysis and produced more lactic acid even in the presence of oxygen. This points out that pyruvate, the end product of glycolysis in most normal cells, is processed one step further in the cytoplasm to generate lactic acid, while in normal physiology it should be transferred to mitochondria. Warburg’s discovery has been confirmed by many reports, as reduced mitochondria content and impaired oxidative phosphorylation are seen in various cancer types^{245, 246} and even in slowly growing tumors²⁴⁷. Below we speculate that some of the alterations observed in our comparison of transformed and non-transformed cells shed light on the molecular basis of the Warburg effect. Note however, that it is difficult to separate differences between cancer cells and primary cells from differences due to adaption to life in cell culture.

Table 4.3.1 Key proteins involved in TCA cycle and glycolysis distribute differently between primary hepatocytes and the transformed hepatocytes Hepa1-6.

| | | |
|------------|---|-----|
| TCA cycle | Fumarate hydratase | 0.2 |
| | Succinate dehydrogenase [ubiquinone] flavoprotein subunit | 0.2 |
| | aconitase | 0.5 |
| Glycolysis | Hexokinase II | 6 |
| | Isoform 1 of 6-phosphofructokinase type C | 4.8 |
| | 6-phosphofructokinase, muscle type | 2.1 |
| | Fructose-biphosphate aldolase A | 3.8 |
| | Pyruvate kinase isozyme M2 | 7.1 |

In our proteome wide comparison study, we quantified key enzymes involved in glucose metabolic processes (Table 4.3.1). Two pivotal mitochondrial enzymes in the TCA cycle that specifically control ATP production²⁴⁸, fumarate hydratase and succinate dehydrogenase, are down-regulated in transformed cells. In contrast, several enzymes involved in glycolysis were over-expressed in Hepa1-6, mainly represented by the key proteins involved in rate-limiting steps such as hexokinase, phosphofructokinase, and pyruvate kinase. In fact, hexokinase II over-expression in hepatoma represented the first significant insight into the molecular basis of “Warburg effect”^{249, 250}. It was revealed that hexokinase II directly binds to the mitochondrial outer membrane and utilizes mitochondrial ATP to rapidly phosphorylate glucose, therefore the first rate-limiting step is overcome and glycolysis is initiated. Tumor cells adopt this strategy to ensure a continuous and rapid flux of glycolysis.

Glycolysis and oxidative phosphorylation are the two major ATP generating pathways within cells. While glycolysis alone can generate two ATP molecules, the complete oxidation of glucose via glycolysis and oxidative phosphorylation can generate much more - 36 molecules ATP. Apparently mitochondria are much more efficient in producing energy. Energy is in constant demand for the rapid proliferation of tumor cells. But what are the reasons for tumor cells to favor the less efficient energy generating pathway?

One significant factor is the shrinkage of mitochondria in neoplasm^{251, 252}. We observed that over 60% of mitochondrial proteins were under-expressed by at least two-fold in the transformed cells (Hepa1-6) compared to non-transformed primary cells. This is in good agreement with our fluorescent staining that mitochondria are drastically reduced (Figure 4.3.9).

Furthermore, accumulated mutations in mitochondrial DNA (mtDNA) often lead to deficient respiration and ATP generation in neoplasm²⁴⁹. One of the causes of mtDNA damage is the side product of oxidative phosphorylation, the reactive oxygen species (ROS). ROS damages several enzymes involved in the TCA cycle, such as aconitase²⁵³. In normal physiology reduction of superoxide is carried out by enzymes. Hydrogen peroxide (H₂O₂) is slowly reduced to water by glutathione peroxidase in mitochondria and by peroxisomal enzyme catalase in the cytosol²⁴⁹. In our study we observed that the mitochondrial enzyme glutathione peroxidase was down-regulated by half in transformed hepatocytes. Therefore it is possible that accumulated ROS contributed to the mitochondrial dysfunction and shrinkage in the transformed cell line Hepa1-6. As a consequence, the impaired mitochondrial functions may contribute to force tumor cells into a switch in the metabolic network.

A third significant factor contributing to the metabolic switch is the oxygen deficiency in the tumor microenvironment. Rapid proliferation drives tumor cells away from blood vessels, and the relatively slow vascularization usually causes a suppressed oxygen supply. As oxygen is the ultimate receiver of electrons in oxidative phosphorylation, this hypoxic environment hinders execution of this cellular respiration process.

While the “Warburg effect” reflects the distortion of homeostasis in the whole tumor system, it actually represents single cell Darwinian evolution. Tumor cells need to develop an alternative strategy to meet the needs of rapid proliferation, growth and survival, and invasion to other tissues. The “Warburg effect” can confer a number of benefits to tumor cells. First, it provides building blocks for synthesizing biomolecules. The elevated glycolysis and its derivative pentose phosphate pathway can generate a large amount of carbon precursors for the synthesis of nucleic acids, phospholipids, fatty acids, cholesterol, porphyrins, and so on. Second, lactic acid production can regenerate NAD⁺ from NADH and therefore equilibrate the initial conditions for glycolysis. Furthermore it creates an acidic environment where tumor cells may be less exposed to attack of the immune system.

As the “Warburg effect” is apparently very beneficial for tumor development, cells use other strategies to promote and strengthen this metabolic switch. Growth factor signaling pathways are often reported to be hyper-active in various tumor types. Besides their canonical functions of enhancing cell proliferation and survival, some of the RTK signaling proteins also influence metabolism. For instance, Ras, PI3K, Akt and BCR-ABL are able to promote glycolysis²⁴⁸. In the absence of oxygen, HIF-1 increases the expression of most glycolytic enzymes and glucose transporters GLUT1 and GLUT3²⁵⁴. Consistent with these previous findings, our study revealed over-expression of several growth factor signaling proteins in cancer cells (Table 4.3.2).

Table 4.3.2 Key proteins involved in cancer development and upregulated in Hepa1-6.

| | Protein Name | Ratio [Hepa1-6/primary] |
|-------------------------|--|------------------------------------|
| Apoptosis | p53 and DNA damage-regulated protein 1 | 3.5 |
| Growth Factor signaling | Phosphatidylinositol 3-kinase regulatory sub alpha | 3 |
| | RAC-alpha serine/threonine-protein kinase (AKT) | 1.5 |
| | Tyrosine-protein kinase ABL2 | 10 |

Enhanced glycolysis is one of the fundamental principles shared by various tumor types. The accumulating knowledge of its molecular basis has led to the development of glycolysis inhibitors for cancer therapy²⁴⁸. The strength of this strategy lies in its applicability to all tumor cells, which bypasses the difficulty of cancer heterogeneity. Heterogeneity is a notorious feature of cancer. Different types of cancer possess multiple genetic and epigenetic alterations. Even within the same type of cancer, malignant cell populations may contain diverse genetic alterations. This genetic instability can evolve over time, leading to ever higher degree of heterogeneity. Therefore target-specific drugs may function only partially in clinics. In contrast, a direct attack on the common weakness of all tumor cells, i.e. a heavy reliance on glycolysis, may simultaneously suppress various tumor types. As healthy cells can consume other energy fuels, such as protein and lipid, glycolysis inhibition may not induce an adverse effect in these cells, reducing potential side effects. However, cautions should still be exercised as there are indeed organs which employ glucose as a major source of energy. These organs include the brain, retina, and testis. For this reason, this inhibition therapy has to be adjusted precisely and evaluated thoroughly.

To date, a few inhibitor compounds targeting glycolysis pathway have been under evaluation²⁴⁸. *2-Deoxyglucose*, a glucose analog competing as hexokinase substrate, and *lonidamine* which inhibits the kinase activity of hexokinase, have both entered clinical trials. Therefore it is likely that in the future, glycolysis inhibition compounds will be applied in clinical therapy and contribute to cancer treatment.

5 Perspectives

Even though the current work flow of quantitative mass spectrometry in cell line studies will continue to contribute to various topics in cancer research, it is foreseeable that future directions will move towards analysis of real clinical samples. Technology innovations will be required to fulfill this task. They can come from three major directions:

1. Innovations in proteomics technique

Clinical samples often come from limited sources. Proteomics technique will have to adjust to decreased sample amounts. Currently, in our laboratory a minimum of ten micrograms of proteins are required for proteome mapping while for phosphoproteome analysis one milligram is generally needed. Therefore there is an urgent need to improve the performance of chromatographic separation and mass spectrometry in analyzing low abundance proteins, especially in post-translational modifications studies.

2. Innovations in biological research models for proteomics studies

For convenience and reproducibility, cell line models are regularly used to study cancer. However, these *ex vivo* systems cannot fully represent *in vivo* conditions. For example, tissue context is lost as single cell populations are cultivated in solid plastic dishes. Oxygen concentration is significantly increased in cell culture incubators compared to normal physiology. As a worse case, hypoxia is a common condition for tumor cell growth, which entails further decreased oxygen concentration. Therefore there is a need to improve the current models for cancer research. For example, studies on primary cells will become more prevalent in the future. Three-dimensional cell culture systems are already frequently used, such as Matrigel. Analysis will be more frequently performed on biopsy samples. In this regard, protein quantitation can be carried out using the corresponding SILAC labeled cell line as an internal standard. In the long run, label-free quantitative mass spectrometry will become more robust and may be the method of choice for quantitative proteomics for patient samples.

3. Innovations in computational biology

Data mining will become even more important with the advent of more large scale datasets. Individual difference should be minimized among patient samples and the significant factors contributing to the disease should be highlighted. Integration of various “omics” data will certainly improve data mining and reveal valuable biological information at the systems level. Mathematical modeling will play an important role, and it may already be widely used in a few years from now.

6 References

1. Lyris M. F. de Godoy¹, Jesper V. Olsen^{1,*}, Jürgen Cox^{1,*}, Michael L. Nielsen^{1,*}, Nina C. Hubner¹, Florian Fröhlich², Tobias C. Walther², and Matthias Mann¹ Comprehensive, mass spectrometry-based proteome quantitation of haploid versus diploid yeast. *Submitted manuscript* (2008).
2. Seger, R., Rodeck, U. & Yarden, Y. Receptor tyrosine kinases: the emerging tip of systems control. *ET Syst Biol* 2, 1-4 (2008).
3. Smithgall, T.E. Signal transduction pathways regulating hematopoietic differentiation. *Pharmacol Rev* 50, 1-19 (1998).
4. Gregoire, F.M., Smas, C.M. & Sul, H.S. Understanding adipocyte differentiation. *Physiol Rev* 78, 783-809 (1998).
5. Vivekanand, P. & Rebay, I. Intersection of signal transduction pathways and development. *Annual review of genetics* 40, 139-157 (2006).
6. Wagner, M. & Siddiqui, M.A. Signal transduction in early heart development (II): ventricular chamber specification, trabeculation, and heart valve formation. *Exp Biol Med (Maywood)* 232, 866-880 (2007).
7. Wagner, M. & Siddiqui, M.A. Signal transduction in early heart development (I): cardiogenic induction and heart tube formation. *Exp Biol Med (Maywood)* 232, 852-865 (2007).
8. Ewbank, J.J. Signaling in the immune response. *WormBook*, 1-12 (2006).
9. Juliano, R.L. Signal transduction by cell adhesion receptors and the cytoskeleton: functions of integrins, cadherins, selectins, and immunoglobulin-superfamily members. *Annual review of pharmacology and toxicology* 42, 283-323 (2002).
10. Talapatra, S. & Thompson, C.B. Growth factor signaling in cell survival: implications for cancer treatment. *J Pharmacol Exp Ther* 298, 873-878 (2001).
11. Garcia-Segura, L.M., Cardona-Gomez, G.P., Chowen, J.A. & Azcoitia, I. Insulin-like growth factor-I receptors and estrogen receptors interact in the promotion of neuronal survival and neuroprotection. *J Neurocytol* 29, 425-437 (2000).
12. Gire, V., Roux, P., Wynford-Thomas, D., Brondello, J.M. & Dulic, V. DNA damage checkpoint kinase Chk2 triggers replicative senescence. *The EMBO journal* 23, 2554-2563 (2004).
13. Swaminathan, G. & Tsygankov, A.Y. The Cbl family proteins: ring leaders in regulation of cell signaling. *Journal of cellular physiology* 209, 21-43 (2006).
14. Tiganis, T. Protein tyrosine phosphatases: dephosphorylating the epidermal growth factor receptor. *IUBMB Life* 53, 3-14 (2002).
15. Pennisi, E. Human genome. A low number wins the GeneSweep Pool. *Science (New York, N.Y)* 300, 1484 (2003).
16. Blume-Jensen, P. & Hunter, T. Oncogenic kinase signalling. *Nature* 411, 355-365 (2001).
17. Venter, J.C. et al. The sequence of the human genome. *Science (New York, N.Y)* 291, 1304-1351 (2001).
18. Manning, G., Whyte, D.B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science (New York, N.Y)* 298, 1912-1934 (2002).
19. Wang, W.Q., Sun, J.P. & Zhang, Z.Y. An overview of the protein tyrosine phosphatase superfamily. *Curr Top Med Chem* 3, 739-748 (2003).
20. Forrest, A.R. et al. Phosphoregulators: protein kinases and protein phosphatases of mouse. *Genome Res* 13, 1443-1454 (2003).
21. Alonso, A. et al. Protein tyrosine phosphatases in the human genome. *Cell* 117, 699-711 (2004).
22. Lander, E.S. et al. Initial sequencing and analysis of the human genome. *Nature* 409, 860-921 (2001).
23. Sammut, S.J., Finn, R.D. & Bateman, A. Pfam 10 years on: 10,000 families and still growing. *Briefings in bioinformatics* 9, 210-219 (2008).

References

24. Bornberg-Bauer, E., Beaussart, F., Kummerfeld, S.K., Teichmann, S.A. & Weiner, J., 3rd The evolution of domain arrangements in proteins and interaction networks. *Cell Mol Life Sci* 62, 435-445 (2005).
25. Nystrom, F.H. & Quon, M.J. Insulin signalling: metabolic pathways and mechanisms for specificity. *Cell Signal* 11, 563-574 (1999).
26. Ogihara, T. et al. Insulin receptor substrate (IRS)-2 is dephosphorylated more rapidly than IRS-1 via its association with phosphatidylinositol 3-kinase in skeletal muscle cells. *The Journal of biological chemistry* 272, 12868-12873 (1997).
27. Blagoev, B., Ong, S.E., Kratchmarova, I. & Mann, M. Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nature biotechnology* 22, 1139-1145 (2004).
28. de Godoy, L.M. et al. Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome biology* 7, R50 (2006).
29. Dengjel, J. et al. Quantitative proteomic assessment of very early cellular signaling events. *Nature biotechnology* 25, 566-568 (2007).
30. Chunaram Choudhary¹, Jesper V. Olsen¹, Christian Brandts², Jürgen Cox¹, Frank Böhmer³, Volker Gerke⁴, Dirk-E. Schmidt-Arras³, Wolfgang E. Berdel⁵, Carsten Müller-Tidow⁵, Matthias Mann^{1,*}, Hubert Serve^{2,*} Mislocalized activation of oncogenic RTKs switches downstream signaling outcomes. *Submitted manuscript* (2008).
31. Olsen, J.V. et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127, 635-648 (2006).
32. Gual, P., Le Marchand-Brustel, Y. & Tanti, J.F. Positive and negative regulation of insulin signaling through IRS-1 phosphorylation. *Biochimie* 87, 99-109 (2005).
33. Cohen, P. The past and future of protein phosphatase research. *Methods in enzymology* 366, xlv-xlix, 1 (2003).
34. Ostman, A., Hellberg, C. & Bohmer, F.D. Protein-tyrosine phosphatases and cancer. *Nature reviews* 6, 307-320 (2006).
35. Hershko, A. & Ciechanover, A. The ubiquitin system. *Annu Rev Biochem* 67, 425-479 (1998).
36. Mukhopadhyay, D. & Riezman, H. Proteasome-independent functions of ubiquitin in endocytosis and signaling. *Science (New York, N.Y)* 315, 201-205 (2007).
37. Shaul, Y.D. & Seger, R. The MEK/ERK cascade: from signaling specificity to diverse functions. *Biochimica et biophysica acta* 1773, 1213-1226 (2007).
38. Fecchi, K., Volonte, D., Hezel, M.P., Schmeck, K. & Galbiati, F. Spatial and temporal regulation of GLUT4 translocation by flotillin-1 and caveolin-3 in skeletal muscle cells. *Faseb J* 20, 705-707 (2006).
39. Saltiel, A.R. & Pessin, J.E. Insulin signaling pathways in time and space. *Trends Cell Biol* 12, 65-71 (2002).
40. Modrek, B. & Lee, C. A genomic view of alternative splicing. *Nature genetics* 30, 13-19 (2002).
41. Thanaraj, T.A. et al. ASD: the Alternative Splicing Database. *Nucleic acids research* 32, D64-69 (2004).
42. Gomperts, B.D.K., I M; Tatham E R Signal Transduction. *Book* (2003).
43. Davidson, E.A. & Hirsch, A.I. Carbon cycle. Fertile forest experiments. *Nature* 411, 431-433 (2001).
44. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* 100, 57-70 (2000).
45. Karin, M. Nuclear factor-kappaB in cancer development and progression. *Nature* 441, 431-436 (2006).
46. Evan, G.I. & Vousden, K.H. Proliferation, cell cycle and apoptosis in cancer. *Nature* 411, 342-348 (2001).
47. Harbour, J.W. & Dean, D.C. The Rb/E2F pathway: expanding roles and emerging paradigms. *Genes Dev* 14, 2393-2409 (2000).
48. Baudino, T.A. & Cleveland, J.L. The Max network gone mad. *Molecular and cellular biology* 21, 691-702 (2001).

References

49. Maddika, S. et al. Cell survival, cell death and cell cycle pathways are interconnected: implications for cancer therapy. *Drug Resist Updat* 10, 13-29 (2007).
50. Shaw, R.J. & Cantley, L.C. Ras, PI(3)K and mTOR signalling controls tumour cell growth. *Nature* 441, 424-430 (2006).
51. Garnett, M.J. & Marais, R. Guilty as charged: B-RAF is a human oncogene. *Cancer cell* 6, 313-319 (2004).
52. Luo, J., Manning, B.D. & Cantley, L.C. Targeting the PI3K-Akt pathway in human cancer: rationale and promise. *Cancer cell* 4, 257-262 (2003).
53. New, D.C. & Wong, Y.H. Molecular mechanisms mediating the G protein-coupled receptor regulation of cell cycle progression. *Journal of molecular signaling* 2, 2 (2007).
54. Cohen, P. Protein kinases--the major drug targets of the twenty-first century? *Nature reviews* 1, 309-315 (2002).
55. Pouyssegur, J., Dayan, F. & Mazure, N.M. Hypoxia signalling in cancer and approaches to enforce tumour regression. *Nature* 441, 437-443 (2006).
56. Yu, H. & Rohan, T. Role of the insulin-like growth factor family in cancer development and progression. *Journal of the National Cancer Institute* 92, 1472-1489 (2000).
57. Datta, S.R., Brunet, A. & Greenberg, M.E. Cellular survival: a play in three Akts. *Genes Dev* 13, 2905-2927 (1999).
58. Stambolic, V., Mak, T.W. & Woodgett, J.R. Modulation of cellular apoptotic potential: contributions to oncogenesis. *Oncogene* 18, 6094-6103 (1999).
59. Maehama, T. & Dixon, J.E. PTEN: a tumour suppressor that functions as a phospholipid phosphatase. *Trends Cell Biol* 9, 125-128 (1999).
60. Albrechtsen, N. et al. Maintenance of genomic integrity by p53: complementary roles for activated and non-activated p53. *Oncogene* 18, 7706-7717 (1999).
61. Christofori, G. New signals from the invasive front. *Nature* 441, 444-450 (2006).
62. Mansuy, I.M. & Shenolikar, S. Protein serine/threonine phosphatases in neuronal plasticity and disorders of learning and memory. *Trends Neurosci* 29, 679-686 (2006).
63. Andreeva, A.V. & Kutuzov, M.A. PPP family of protein Ser/Thr phosphatases: two distinct branches? *Molecular biology and evolution* 18, 448-452 (2001).
64. Khoury, J.D., Rassidakis, G.Z., Medeiros, L.J., Amin, H.M. & Lai, R. Methylation of SHP1 gene and loss of SHP1 protein expression are frequent in systemic anaplastic large cell lymphoma. *Blood* 104, 1580-1581 (2004).
65. Oka, T. et al. Gene silencing of the tyrosine phosphatase SHP1 gene by aberrant methylation in leukemias/lymphomas. *Cancer research* 62, 6390-6394 (2002).
66. Chim, C.S., Fung, T.K., Cheung, W.C., Liang, R. & Kwong, Y.L. SOCS1 and SHP1 hypermethylation in multiple myeloma: implications for epigenetic activation of the Jak/STAT pathway. *Blood* 103, 4630-4635 (2004).
67. Motiwala, T. et al. Protein tyrosine phosphatase receptor-type O (PTPRO) exhibits characteristics of a candidate tumor suppressor in human lung cancer. *Proceedings of the National Academy of Sciences of the United States of America* 101, 13844-13849 (2004).
68. Tartaglia, M. et al. Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. *Nature genetics* 29, 465-468 (2001).
69. Tartaglia, M. et al. Genetic evidence for lineage-related and differentiation stage-related contribution of somatic PTPN11 mutations to leukemogenesis in childhood acute leukemia. *Blood* 104, 307-313 (2004).
70. Liu, J. et al. Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes. *Cell* 66, 807-815 (1991).
71. Davies, S.P., Reddy, H., Caivano, M. & Cohen, P. Specificity and mechanism of action of some commonly used protein kinase inhibitors. *The Biochemical journal* 351, 95-105 (2000).
72. Hoessel, R. et al. Indirubin, the active constituent of a Chinese antileukaemia medicine, inhibits cyclin-dependent kinases. *Nature cell biology* 1, 60-67 (1999).
73. Druker, B.J. Molecularly targeted therapy: have the floodgates opened? *The oncologist* 9, 357-360 (2004).

74. Schindler, T. et al. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science (New York, N.Y)* 289, 1938-1942 (2000).
75. Petrelli, A. & Giordano, S. From single- to multi-target drugs in cancer therapy: when aspecificity becomes an advantage. *Current medicinal chemistry* 15, 422-432 (2008).
76. Merlano, M. & Occelli, M. Review of cetuximab in the treatment of squamous cell carcinoma of the head and neck. *Therapeutics and clinical risk management* 3, 871-876 (2007).
77. Semenza, G.L. A new weapon for attacking tumor blood vessels. *The New England journal of medicine* 358, 2066-2067 (2008).
78. Daub, H., Specht, K. & Ullrich, A. Strategies to overcome resistance to targeted protein kinase inhibitors. *Nature reviews* 3, 1001-1010 (2004).
79. Pao, W. et al. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med* 2, e73 (2005).
80. Capdeville, R., Buchdunger, E., Zimmermann, J. & Matter, A. Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nature reviews* 1, 493-502 (2002).
81. Engelman, J.A. et al. MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. *Science (New York, N.Y)* 316, 1039-1043 (2007).
82. Stommel, J.M. et al. Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies. *Science (New York, N.Y)* 318, 287-290 (2007).
83. Shah, N.P. et al. Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. *Cancer cell* 2, 117-125 (2002).
84. Shah, N.P. et al. Sequential ABL kinase inhibitor therapy selects for compound drug-resistant BCR-ABL mutations with altered oncogenic potency. *The Journal of clinical investigation* 117, 2562-2569 (2007).
85. Talpaz, M. et al. Dasatinib in imatinib-resistant Philadelphia chromosome-positive leukemias. *The New England journal of medicine* 354, 2531-2541 (2006).
86. Cravatt, B.F., Simon, G.M. & Yates, J.R., 3rd The biological impact of mass-spectrometry-based proteomics. *Nature* 450, 991-1000 (2007).
87. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* 422, 198-207 (2003).
88. Waanders, L.F., Hanke, S. & Mann, M. Top-down quantitation and characterization of SILAC-labeled proteins. *Journal of the American Society for Mass Spectrometry* 18, 2058-2064 (2007).
89. Graumann, J. et al. Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. *Mol Cell Proteomics* 7, 672-683 (2008).
90. Liu, H., Lin, D. & Yates, J.R., 3rd Multidimensional separations for protein/peptide analysis in the post-genomic era. *Biotechniques* 32, 898, 900, 902 passim (2002).
91. Motoyama, A., Xu, T., Ruse, C.I., Wohlschlegel, J.A. & Yates, J.R., 3rd Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of peptides and phosphopeptides. *Analytical chemistry* 79, 3623-3634 (2007).
92. Aiping Lu, J.R.W., and Matthias Mann Comparative Proteomic Profiling of Membrane Proteins in Rat Cerebellum, Spinal Cord, and Sciatic Nerve. *Manuscript to submit* (2008).
93. Blagoev, B. et al. A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nature biotechnology* 21, 315-318 (2003).
94. Ficarro, S.B. et al. Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nature biotechnology* 20, 301-305 (2002).
95. Gruhler, A. et al. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol Cell Proteomics* 4, 310-327 (2005).
96. Larsen, M.R., Thingholm, T.E., Jensen, O.N., Roepstorff, P. & Jorgensen, T.J. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol Cell Proteomics* 4, 873-886 (2005).

97. Thingholm, T.E., Jensen, O.N., Robinson, P.J. & Larsen, M.R. SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. *Mol Cell Proteomics* 7, 661-671 (2008).
98. Aiping Lu, L.F.W., reinaldo Almeida, guoqing Li, Mark Allen, Juergen Cox, Jesper V Olsen, Tiziana Bonaldi and Matthias Mann Nanoelectrospray peptide mapping revisited: Composite survey spectra allow high dynamic range protein characterization without LCMS on an orbitrap mass spectrometer *International Journal of Mass Spectrometry* 268 (2007).
99. Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S.H. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* 132, 1049-1061 (2008).
100. Uetz, P. et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623-627 (2000).
101. Gavin, A.C. et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141-147 (2002).
102. Michael Karas, U.B., Ulrich Gießmann Matrix-assisted laser desorption ionization mass spectrometry. *Mass Spectrometry Reviews* 10, 335 (1991).
103. John B. Fenn, M.M., Chin Kai Meng, Shek Fu Wong, Craig M. Whitehouse Electrospray ionization-principles and practice. *Mass Spectrometry Reviews* 9, 37-70 (1990).
104. Herzog, J.M.a.R. Mass spectrograph. *Z. Physik* 89 (1934).
105. Bristow, A.W. Accurate mass measurement for the determination of elemental formula--a tutorial. *Mass spectrometry reviews* 25, 99-111 (2006).
106. Nier, E.G.J.a.A.O. Angular Aberrations in Sector Shaped Electromagnetic Lenses for Focusing Beams of Charged Particles. *Phys. Rev.* 91 (1953).
107. Guilhaus, M. Principles and instrumentation in time-of-flight mass spectrometry. Physical and instrumental concepts. *Journal of Mass Spectrometry* 30 (2005).
108. Douglas, D.J., Frank, A.J. & Mao, D. Linear ion traps in mass spectrometry. *Mass spectrometry reviews* 24, 1-29 (2005).
109. March, R.E. An introduction to quadrupole ion trap mass spectrometry. *Journal of Mass Spectrometry* 32 (1997).
110. Amster, I.J. Fourier Transform Mass Spectrometry. *Journal of Mass Spectrometry* 31 (1996).
111. Marshall, A.G., Hendrickson, C.L. & Jackson, G.S. Fourier transform ion cyclotron resonance mass spectrometry: a primer. *Mass spectrometry reviews* 17, 1-35 (1998).
112. Makarov, A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Analytical chemistry* 72, 1156-1162 (2000).
113. McLuckey, S.A. & Wells, J.M. Mass analysis at the advent of the 21st century. *Chemical reviews* 101, 571-606 (2001).
114. Malcolm Dole, L.L.M., and R. L. Hines Molecular Beams of Macroions. *J. Chem. Phys.* 49 (1968).
115. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F. & Whitehouse, C.M. Electrospray ionization for mass spectrometry of large biomolecules. *Science (New York, N.Y)* 246, 64-71 (1989).
116. Mann, M.S.W.a.M. Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? *International Journal of Mass Spectrometry and Ion Processes* 136 (1994).
117. Wilm, M. & Mann, M. Analytical properties of the nanoelectrospray ion source. *Analytical chemistry* 68, 1-8 (1996).
118. NIELSEN, M.L. Characterization of Polypeptides by Tandem Mass Spectrometry Using Complementary Fragmentation Techniques. *Ph.D. Thesis* (2007).
119. Mora, J.F. et al. Electrochemical processes in electrospray ionization mass spectrometry. *J Mass Spectrom* 35, 939-952 (2000).
120. Schwartz, J.C., Senko, M.W. & Syka, J.E. A two-dimensional quadrupole ion trap mass spectrometer. *Journal of the American Society for Mass Spectrometry* 13, 659-669 (2002).
121. A.G., W.Y.S.S.D.-H.H.C.L.M. Mass-selective ion accumulation and fragmentation in a linear octopole ion trap external to a fourier transform ion cyclotron resonance mass spectrometer. *International Journal of Mass Spectrometry* 198 (2000).

122. John N. Louris, R.G.C., John E. P. Syka, Paul E. Kelley, George C. Stafford, and John F. J. Todd Instrumentation, applications, and energy deposition in quadrupole ion-trap tandem mass spectrometry. *Anal. Chem.* 59 (1987).
123. Mayya, V., Rezaul, K., Cong, Y.S. & Han, D. Systematic comparison of a two-dimensional ion trap and a three-dimensional ion trap mass spectrometer in proteomics. *Mol Cell Proteomics* 4, 214-223 (2005).
124. Adele R. Blackler, A.A.K., ‡ Michael J. MacCoss,‡ and Christine C. Wu*,† Quantitative Comparison of Proteomic Data Quality between a 2D and 3D Quadrupole Ion Trap. *Anal. Chem.* 78 (2006).
125. Gorshkov, M.V. & Zubarev, R.A. On the accuracy of polypeptide masses measured in a linear ion trap. *Rapid Commun Mass Spectrom* 19, 3755-3758 (2005).
126. Li, Y., Hunter, R.L. & Mclver, R.T., Jr. High-resolution mass spectrometer for protein chemistry. *Nature* 370, 393-395 (1994).
127. Hu, Q. et al. The Orbitrap: a new mass spectrometer. *J Mass Spectrom* 40, 430-443 (2005).
128. Hardman, M. & Makarov, A.A. Interfacing the orbitrap mass analyzer to an electrospray ion source. *Analytical chemistry* 75, 1699-1705 (2003).
129. Scigelova, M. & Makarov, A. Orbitrap mass analyzer--overview and applications in proteomics. *Proteomics* 6 Suppl 2, 16-21 (2006).
130. Olsen, J.V. et al. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* 4, 2010-2021 (2005).
131. Syka, J.E. et al. Novel linear quadrupole ion trap/FT mass spectrometer: performance characterization and use in the comparative analysis of histone H3 post-translational modifications. *Journal of proteome research* 3, 621-626 (2004).
132. Yates, J.R., Cociorva, D., Liao, L. & Zabrouskov, V. Performance of a linear ion trap-Orbitrap hybrid for peptide analysis. *Analytical chemistry* 78, 493-500 (2006).
133. P. Roepstorff 1, J.F. Proposal for a Common Nomenclature for sequence Ions in Mass Spectra of Peptides. (1984).
134. Johnson, R.S., Martin, S.A., Biemann, K., Stults, J.T. & Watson, J.T. Novel fragmentation process of peptides by collision-induced decomposition in a tandem mass spectrometer: differentiation of leucine and isoleucine. *Analytical chemistry* 59, 2621-2625 (1987).
135. Olsen, J.V. & Mann, M. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proceedings of the National Academy of Sciences of the United States of America* 101, 13417-13422 (2004).
136. Savitski, M.M., Kjeldsen, F., Nielsen, M.L. & Zubarev, R.A. Complementary sequence preferences of electron-capture dissociation and vibrational excitation in fragmentation of polypeptide polycations. *Angewandte Chemie (International ed)* 45, 5301-5303 (2006).
137. Zhang, R. & Regnier, F.E. Minimizing resolution of isotopically coded peptides in comparative proteomics. *Journal of proteome research* 1, 139-147 (2002).
138. Ong, S.E. et al. 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 (2002).
139. Zhang, R., Sioma, C.S., Wang, S. & Regnier, F.E. Fractionation of isotopically labeled peptides in quantitative proteomics. *Analytical chemistry* 73, 5142-5149 (2001).
140. Mann, J.C.a.M. High peptide identification rates and proteome-wide quantitation via novel computational strategies. *submitted manuscript* (2008).
141. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Analytical and bioanalytical chemistry* 389, 1017-1031 (2007).
142. Ong, S.E. & Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nature chemical biology* 1, 252-262 (2005).
143. Oda, Y., Huang, K., Cross, F.R., Cowburn, D. & Chait, B.T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America* 96, 6591-6596 (1999).

144. Krijgsveld, J. et al. Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics. *Nature biotechnology* 21, 927-931 (2003).
145. Wu, C.C., MacCoss, M.J., Howell, K.E., Matthews, D.E. & Yates, J.R., 3rd Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Analytical chemistry* 76, 4951-4959 (2004).
146. Marcus Krueger¹, Markus Moser^{2,3}, Siegfried Ussar², Ingo Thievensen², Christian A. Luber¹, Francesca Forner¹, Sarah Schmidt², Sara Zanivan¹, Reinhard Fässler², and Matthias Mann¹ SILAC-mouse for quantitative proteomics uncovers Kindlin-3 as an essential factor for red blood cell function. *Cell* accepted manuscript (2008).
147. Gygi, S.P. et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature biotechnology* 17, 994-999 (1999).
148. Ross, P.L. et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 3, 1154-1169 (2004).
149. Yao, X., Freas, A., Ramirez, J., Demirev, P.A. & Fenselau, C. Proteolytic ¹⁸O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Analytical chemistry* 73, 2836-2842 (2001).
150. Reynolds, K.J., Yao, X. & Fenselau, C. Proteolytic ¹⁸O labeling for comparative proteomics: evaluation of endoprotease Glu-C as the catalytic agent. *Journal of proteome research* 1, 27-33 (2002).
151. Mann, M. Functional and quantitative proteomics using SILAC. *Nature reviews* 7, 952-958 (2006).
152. Pierce, A. et al. Eight-channel iTRAQ enables comparison of the activity of six leukemogenic tyrosine kinases. *Mol Cell Proteomics* 7, 853-863 (2008).
153. Hanke, S., Besir, H., Oesterhelt, D. & Mann, M. Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level. *Journal of proteome research* 7, 1118-1130 (2008).
154. Reinders, J. & Sickmann, A. State-of-the-art in phosphoproteomics. *Proteomics* 5, 4052-4061 (2005).
155. Schmelzle, K., Kane, S., Gridley, S., Lienhard, G.E. & White, F.M. Temporal dynamics of tyrosine phosphorylation in insulin signaling. *Diabetes* 55, 2171-2179 (2006).
156. Michelle T. Barati, M.J.R., Jon B. Klein, and Kenneth R. McLeish A Proteomic Screen Identified Stress-Induced Chaperone Proteins as Targets of Akt Phosphorylation in Mesangial Cells *J. Proteome Res.* 5 (2006).
157. Andersson, L. & Porath, J. Isolation of phosphoproteins by immobilized metal (Fe³⁺) affinity chromatography. *Analytical biochemistry* 154, 250-254 (1986).
158. Posewitz, M.C. & Tempst, P. Immobilized gallium(III) affinity chromatography of phosphopeptides. *Analytical chemistry* 71, 2883-2892 (1999).
159. Ballif, B.A., Villen, J., Beausoleil, S.A., Schwartz, D. & Gygi, S.P. Phosphoproteomic analysis of the developing mouse brain. *Mol Cell Proteomics* 3, 1093-1101 (2004).
160. Oda, Y., Nagasu, T. & Chait, B.T. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nature biotechnology* 19, 379-382 (2001).
161. Zhou, H., Watts, J.D. & Aebersold, R. A systematic approach to the analysis of protein phosphorylation. *Nature biotechnology* 19, 375-378 (2001).
162. Amoresano, A., Marino, G., Cirulli, C. & Quemeneur, E. Mapping phosphorylation sites: a new strategy based on the use of isotopically labelled DTT and mass spectrometry. *European journal of mass spectrometry (Chichester, England)* 10, 401-412 (2004).
163. Vosseller, K. et al. Quantitative analysis of both protein expression and serine / threonine post-translational modifications through stable isotope labeling with dithiothreitol. *Proteomics* 5, 388-398 (2005).
164. F Rusnak, J.Z., and GM Hathaway Identification of phosphorylated and glycosylated sites in peptides by chemically targeted proteolysis *Journal of Biomolecular Techniques* 13 (2002).
165. Knight, Z.A. et al. Phosphospecific proteolysis for mapping sites of protein phosphorylation. *Nature biotechnology* 21, 1047-1054 (2003).

166. Beausoleil, S.A. et al. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proceedings of the National Academy of Sciences of the United States of America* 101, 12130-12135 (2004).
167. Masahiro Kawahara, H.N.a.T.N. Titania and zirconia: possible new ceramic microparticulates for high-performance liquid chromatography. *Journal of Chromatography A* 515 (1990).
168. Pinkse, M.W., Uitto, P.M., Hilhorst, M.J., Ooms, B. & Heck, A.J. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Analytical chemistry* 76, 3935-3943 (2004).
169. Attwood, P.V., Piggott, M.J., Zu, X.L. & Besant, P.G. Focus on phosphohistidine. *Amino acids* 32, 145-156 (2007).
170. Darlington, G.J. Liver cell lines. *Methods in enzymology* 151, 19-38 (1987).
171. Liu, Y. et al. Enhancement of immunogenicity of tumor cells by cotransfection with genes encoding antisense insulin-like growth factor-1 and B7.1 molecules. *Cancer gene therapy* 7, 456-465 (2000).
172. Nagayama, Y., Nakao, K., Mizuguchi, H., Hayakawa, T. & Niwa, M. Enhanced antitumor effect of combined replicative adenovirus and nonreplicative adenovirus expressing interleukin-12 in an immunocompetent mouse model. *Gene therapy* 10, 1400-1403 (2003).
173. Klein, E. et al. Properties of the K562 cell line, derived from a patient with chronic myeloid leukemia. *International journal of cancer* 18, 421-431 (1976).
174. Andersson, L.C., Nilsson, K. & Gahmberg, C.G. K562--a human erythroleukemic cell line. *International journal of cancer* 23, 143-147 (1979).
175. Klingmuller, U. et al. Primary mouse hepatocytes for systems biology approaches: a standardized in vitro system for modelling of signal transduction pathways. *Systems biology* 153, 433-447 (2006).
176. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Analytical chemistry* 75, 663-670 (2003).
177. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V. & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature protocols* 1, 2856-2860 (2006).
178. Olsen, J.V., Ong, S.E. & Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics* 3, 608-614 (2004).
179. Macek, B. et al. The serine/threonine/tyrosine phosphoproteome of the model bacterium *Bacillus subtilis*. *Mol Cell Proteomics* 6, 697-707 (2007).
180. Schroeder, M.J., Shabanowitz, J., Schwartz, J.C., Hunt, D.F. & Coon, J.J. A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. *Analytical chemistry* 76, 3590-3598 (2004).
181. Gnad, F. et al. PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome biology* 8, R250 (2007).
182. Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics (Oxford, England)* 21, 3448-3449 (2005).
183. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics (Oxford, England)* 23, 257-258 (2007).
184. Team, R.D.C. R: a language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna, Austria; 2004).
185. Sturn, A., Quackenbush, J. & Trajanoski, Z. Genesis: cluster analysis of microarray data. *Bioinformatics (Oxford, England)* 18, 207-208 (2002).
186. Schwartz, D. & Gygi, S.P. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nature biotechnology* 23, 1391-1398 (2005).
187. Huyer, G. et al. Mechanism of inhibition of protein-tyrosine phosphatases by vanadate and pervanadate. *The Journal of biological chemistry* 272, 843-851 (1997).

188. Resjo, S., Oknianska, A., Zolnierowicz, S., Manganiello, V. & Degerman, E. Phosphorylation and activation of phosphodiesterase type 3B (PDE3B) in adipocytes in response to serine/threonine phosphatase inhibitors: deactivation of PDE3B in vitro by protein phosphatase type 2A. *The Biochemical journal* 341 (Pt 3), 839-845 (1999).
189. Ishihara, H. et al. Calyculin A and okadaic acid: inhibitors of protein phosphatase activity. *Biochemical and biophysical research communications* 159, 871-877 (1989).
190. Enan, E. & Matsumura, F. Specific inhibition of calcineurin by type II synthetic pyrethroid insecticides. *Biochemical pharmacology* 43, 1777-1784 (1992).
191. Villen, J., Beausoleil, S.A., Gerber, S.A. & Gygi, S.P. Large-scale phosphorylation analysis of mouse liver. *Proceedings of the National Academy of Sciences of the United States of America* 104, 1488-1493 (2007).
192. Makarov, A., Denisov, E., Lange, O. & Horning, S. Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. *Journal of the American Society for Mass Spectrometry* 17, 977-982 (2006).
193. Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W. & Gygi, S.P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proceedings of the National Academy of Sciences of the United States of America* 100, 6940-6945 (2003).
194. Steen, H., Jebanathirajah, J.A., Springer, M. & Kirschner, M.W. Stable isotope-free relative and absolute quantitation of protein phosphorylation stoichiometry by MS. *Proceedings of the National Academy of Sciences of the United States of America* 102, 3948-3953 (2005).
195. Boeri Erba, E. et al. Quantitation of multisite EGF receptor phosphorylation using mass spectrometry and a novel normalization approach. *Journal of proteome research* 6, 2768-2785 (2007).
196. Knight, Z.A. & Shokat, K.M. Features of selective kinase inhibitors. *Chemistry & biology* 12, 621-637 (2005).
197. Bain, J., McLauchlan, H., Elliott, M. & Cohen, P. The specificities of protein kinase inhibitors: an update. *The Biochemical journal* 371, 199-204 (2003).
198. Fabian, M.A. et al. A small molecule-kinase interaction map for clinical kinase inhibitors. *Nature biotechnology* 23, 329-336 (2005).
199. Becker, F. et al. A three-hybrid approach to scanning the proteome for targets of small molecule kinase inhibitors. *Chemistry & biology* 11, 211-223 (2004).
200. Daub, H. Characterisation of kinase-selective inhibitors by chemical proteomics. *Biochimica et biophysica acta* 1754, 183-190 (2005).
201. Bantscheff, M. et al. Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nature biotechnology* 25, 1035-1044 (2007).
202. Valet, G. Cytomics as a new potential for drug discovery. *Drug discovery today* 11, 785-791 (2006).
203. Nicholson, R.L., Welch, M., Ladlow, M. & Spring, D.R. Small-molecule screening: advances in microarraying and cell-imaging technologies. *ACS chemical biology* 2, 24-30 (2007).
204. Lang, P., Yeow, K., Nichols, A. & Scheer, A. Cellular imaging in drug discovery. *Nature reviews* 5, 343-356 (2006).
205. Krutzik, P.O., Crane, J.M., Clutter, M.R. & Nolan, G.P. High-content single-cell drug screening with phosphospecific flow cytometry. *Nature chemical biology* 4, 132-142 (2008).
206. Bose, R. et al. Phosphoproteomic analysis of Her2/neu signaling and inhibition. *Proceedings of the National Academy of Sciences of the United States of America* 103, 9773-9778 (2006).
207. Tong, L. et al. A highly specific inhibitor of human p38 MAP kinase binds in the ATP pocket. *Nature structural biology* 4, 311-316 (1997).
208. Nishimoto, S. & Nishida, E. MAPK signalling: ERK5 versus ERK1/2. *EMBO reports* 7, 782-786 (2006).
209. Johnson, G.L. & Lapadat, R. Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science (New York, N.Y)* 298, 1911-1912 (2002).
210. Morton, S., Davis, R.J. & Cohen, P. Signalling pathways involved in multisite phosphorylation of the transcription factor ATF-2. *FEBS letters* 572, 177-183 (2004).

211. Smith, J.A., Poteet-Smith, C.E., Malarkey, K. & Sturgill, T.W. Identification of an extracellular signal-regulated kinase (ERK) docking site in ribosomal S6 kinase, a sequence critical for activation by ERK in vivo. *The Journal of biological chemistry* 274, 2893-2898 (1999).
212. Gazel, A., Nijhawan, R.I., Walsh, R. & Blumenberg, M. Transcriptional profiling defines the roles of ERK and p38 kinases in epidermal keratinocytes. *Journal of cellular physiology* 215, 292-308 (2008).
213. Marais, R., Light, Y., Paterson, H.F., Mason, C.S. & Marshall, C.J. Differential regulation of Raf-1, A-Raf, and B-Raf by oncogenic ras and tyrosine kinases. *The Journal of biological chemistry* 272, 4378-4383 (1997).
214. Rouse, J. et al. A novel kinase cascade triggered by stress and heat shock that stimulates MAPKAP kinase-2 and phosphorylation of the small heat shock proteins. *Cell* 78, 1027-1037 (1994).
215. Nowell, P.C. & Hungerford, D.A. Chromosome studies on normal and leukemic human leukocytes. *Journal of the National Cancer Institute* 25, 85-109 (1960).
216. Goss, V.L. et al. A common phosphotyrosine signature for the Bcr-Abl kinase. *Blood* 107, 4888-4897 (2006).
217. Chabner, B.A. The oncologic four-minute mile. *The oncologist* 6, 230-232 (2001).
218. Steinberg, M. Dasatinib: a tyrosine kinase inhibitor for the treatment of chronic myelogenous leukemia and philadelphia chromosome-positive acute lymphoblastic leukemia. *Clinical therapeutics* 29, 2289-2308 (2007).
219. Shah, N.P. et al. Overriding imatinib resistance with a novel ABL kinase inhibitor. *Science (New York, N.Y)* 305, 399-401 (2004).
220. O'Hare, T. et al. In vitro activity of Bcr-Abl inhibitors AMN107 and BMS-354825 against clinically relevant imatinib-resistant Abl kinase domain mutants. *Cancer research* 65, 4500-4505 (2005).
221. Lombardo, L.J. et al. Discovery of N-(2-chloro-6-methyl-phenyl)-2-(6-(4-(2-hydroxyethyl)-piperazin-1-yl)-2-methylpyrimidin-4-ylamino)thiazole-5-carboxamide (BMS-354825), a dual Src/Abl kinase inhibitor with potent antitumor activity in preclinical assays. *Journal of medicinal chemistry* 47, 6658-6661 (2004).
222. Hantschel, O. et al. The Btk tyrosine kinase is a major target of the Bcr-Abl inhibitor dasatinib. *Proceedings of the National Academy of Sciences of the United States of America* 104, 13283-13288 (2007).
223. Gupta, S., Seth, A. & Davis, R.J. Transactivation of gene expression by Myc is inhibited by mutation at the phosphorylation sites Thr-58 and Ser-62. *Proceedings of the National Academy of Sciences of the United States of America* 90, 3216-3220 (1993).
224. Gregory, M.A., Qi, Y. & Hann, S.R. Phosphorylation by glycogen synthase kinase-3 controls c-myc proteolysis and subnuclear localization. *The Journal of biological chemistry* 278, 51606-51612 (2003).
225. Yada M, H.S., Kamura T, Nishiyama M, Tsunematsu R, Imaki H, Ishida N, Okumura F, Nakayama K, Nakayama KI. Phosphorylation-dependent degradation of c-Myc is mediated by the F-box protein Fbw7. *The EMBO journal* 23 (2004).
226. Weisberg, E., Manley, P.W., Cowan-Jacob, S.W., Hochhaus, A. & Griffin, J.D. Second generation inhibitors of BCR-ABL for the treatment of imatinib-resistant chronic myeloid leukaemia. *Nature reviews* 7, 345-356 (2007).
227. Jagani, Z., Singh, A. & Khosravi-Far, R. FoxO tumor suppressors and BCR-ABL-induced leukemia: a matter of evasion of apoptosis. *Biochimica et biophysica acta* 1785, 63-84 (2008).
228. Emanuel, P.D., Bates, L.J., Castleberry, R.P., Gualtieri, R.J. & Zuckerman, K.S. Selective hypersensitivity to granulocyte-macrophage colony-stimulating factor by juvenile chronic myeloid leukemia hematopoietic progenitors. *Blood* 77, 925-929 (1991).
229. Bedi, A., Zehnbaauer, B.A., Barber, J.P., Sharkis, S.J. & Jones, R.J. Inhibition of apoptosis by BCR-ABL in chronic myeloid leukemia. *Blood* 83, 2038-2044 (1994).
230. McGahon, A. et al. BCR-ABL maintains resistance of chronic myelogenous leukemia cells to apoptotic cell death. *Blood* 83, 1179-1187 (1994).

231. Carter, T.A. et al. Inhibition of drug-resistant mutants of ABL, KIT, and EGF receptor kinases. *Proceedings of the National Academy of Sciences of the United States of America* 102, 11011-11016 (2005).
232. Faivre, S., Demetri, G., Sargent, W. & Raymond, E. Molecular basis for sunitinib efficacy and future clinical development. *Nature reviews* 6, 734-745 (2007).
233. Klejman, A., Rushen, L., Morrione, A., Slupianek, A. & Skorski, T. Phosphatidylinositol-3 kinase inhibitors enhance the anti-leukemia effect of STI571. *Oncogene* 21, 5868-5876 (2002).
234. Ashburner, M. & Lewis, S. On ontologies for biologists: the Gene Ontology--untangling the web. *Novartis Foundation symposium* 247, 66-80; discussion 80-63, 84-90, 244-252 (2002).
235. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic acids research* 32, D277-280 (2004).
236. Jakowlew, S.B. Transforming growth factor-beta in cancer and metastasis. *Cancer metastasis reviews* 25, 435-457 (2006).
237. Aitken, A.E., Richardson, T.A. & Morgan, E.T. Regulation of drug-metabolizing enzymes and transporters in inflammation. *Annual review of pharmacology and toxicology* 46, 123-149 (2006).
238. Kawajiri, K. & Fujii-Kuriyama, Y. Cytochrome P450 gene regulation and physiological functions mediated by the aryl hydrocarbon receptor. *Archives of biochemistry and biophysics* 464, 207-212 (2007).
239. Rivera, S.P., Saarikoski, S.T. & Hankinson, O. Identification of a novel dioxin-inducible cytochrome P450. *Molecular pharmacology* 61, 255-259 (2002).
240. Hewitt, N.J. et al. Primary hepatocytes: current understanding of the regulation of metabolic enzymes and transporter proteins, and pharmaceutical practice for the use of hepatocytes in metabolism, enzyme induction, transporter, clearance, and hepatotoxicity studies. *Drug metabolism reviews* 39, 159-234 (2007).
241. Guidotti, J.E. et al. Liver cell polyploidization: a pivotal role for binuclear hepatocytes. *The Journal of biological chemistry* 278, 19095-19101 (2003).
242. Warburg, O., Posener, K. & Negelein, E. Ueber den Stoffwechsel der Tumoren. *Biochemische Zeitschrift* 152, 319-344 (1930).
243. Cox, J. & Mann, M. Is proteomics the new genomics? *Cell* 130, 395-398 (2007).
244. Butte, A. The use and analysis of microarray data. *Nature reviews* 1, 951-960 (2002).
245. Pedersen, P.L. Tumor mitochondria and the bioenergetics of cancer cells. *Progress in experimental tumor research. Fortschritte der experimentellen Tumorforschung* 22, 190-274 (1978).
246. Pedersen, P.L. Warburg, me and Hexokinase 2: Multiple discoveries of key molecular events underlying one of cancers' most common phenotypes, the "Warburg Effect", i.e., elevated glycolysis in the presence of oxygen. *Journal of bioenergetics and biomembranes* 39, 211-222 (2007).
247. Burk, D., Woods, M. & Hunter, J. On the significance of glucoysis for cancer growth, with special reference to Morris rat hepatomas. *Journal of the National Cancer Institute* 38, 839-863 (1967).
248. Pelicano, H., Martin, D.S., Xu, R.H. & Huang, P. Glycolysis inhibition for anticancer treatment. *Oncogene* 25, 4633-4646 (2006).
249. Brandon, M., Baldi, P. & Wallace, D.C. Mitochondrial mutations in cancer. *Oncogene* 25, 4647-4662 (2006).
250. Bustamante, E. & Pedersen, P.L. High aerobic glycolysis of rat hepatoma cells in culture: role of mitochondrial hexokinase. *Proceedings of the National Academy of Sciences of the United States of America* 74, 3735-3739 (1977).
251. Kroemer, G. Mitochondria in cancer. *Oncogene* 25, 4630-4632 (2006).
252. Wallace, D.C. A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annual review of genetics* 39, 359-407 (2005).

References

253. Zhang, S.J. et al. Activation of aconitase in mouse fast-twitch skeletal muscle during contraction-mediated oxidative stress. *American journal of physiology* 293, C1154-1159 (2007).
254. King, A., Selak, M.A. & Gottlieb, E. Succinate dehydrogenase and fumarate hydratase: linking mitochondrial dysfunction and cancer. *Oncogene* 25, 4675-4682 (2006).

Appendix

Appendix 2 χ^2 -test to assess enriched kinase motifs in the Hepal-6 phosphoproteome from project 1.

| motif | kinase | class 1 (observed) | class 1 (expected) | class 1 (chi-square) |
|------------------------|----------|--------------------|--------------------|----------------------|
| R.p[ST] | PKA | 414 | 236.5 | 141.21 |
| R[RK].p[ST] | PKA | 220 | 27 | 1388.57 |
| KR..p[ST] | PKA | 82 | 13.6 | 345.14 |
| S..p[ST] | CK1 | 502 | 357.2 | 64.19 |
| [ST]...pS | CK1 | 716 | 523.8 | 82.02 |
| p[ST]..E | CK2 | 772 | 290.4 | 858.34 |
| pS...S | GSK3 | 502 | 319.7 | 113.67 |
| p[ST]P.[RK] | CDK2 | 265 | 29.5 | 1893.37 |
| R..p[ST] | CAMK2 | 817 | 236.5 | 1510.35 |
| R..p[ST]V | CAMK2 | 41 | 14.3 | 50.02 |
| P.p[ST]P | ERK | 261 | 16 | 3765.95 |
| V.p[ST]P | ERK | 59 | 15.6 | 121.19 |
| PEp[ST]P | ERK | 11 | 1.1 | 89.12 |
| R[RST].p[ST].[ST] | AKT | 134 | 6.5 | 2504.87 |
| R.R..p[ST] | AKT | 164 | 13.4 | 1698 |
| R..p[ST].R | PKC | 7 | 13.4 | 3.07 |
| [LVI].[RK].p[ST] | PKD | 236 | 96.9 | 204.42 |
| [IEV]pY[EG][EDPN][IVL] | LCK | 1 | 0.1 | 8.11 |
| [IVL]pY..[PF] | ABL | 10 | 1.7 | 41.35 |
| [ED]..pY..[DEAGST] | SRC | 5 | 3.8 | 0.4 |
| pY..[ILVM] | ALK | 26 | 18.7 | 3.68 |
| [DPSAEN].pY[VLDEINP] | EGFR | 20 | 12.8 | 4.79 |
| p[ST]P.[KR] | CDK1 | 265 | 29.5 | 1893.37 |
| p[ST]P[KR] | CDK1 | 166 | 29.5 | 636.09 |
| [RK].p[ST][ILV] | Aurora | 133 | 96.9 | 13.77 |
| [RKN]R.p[ST][MILV] | Aurora-A | 61 | 8 | 351.8 |
| [DE].p[ST][VILM].[DE] | PLK | 13 | 12.9 | 0 |
| [ED].p[ST][FLIYWVM] | PLK1 | 73 | 147.1 | 38.69 |
| L..p[ST] | NEK6 | 272 | 412.8 | 53.29 |
| L.R..p[ST] | CHK1/2 | 108 | 23.4 | 307.58 |
| [MILV].[RK]..p[ST] | CHK1 | 260 | 107.4 | 222.54 |
| F..Fp[ST][FY] | PDK1 | 2 | 0.4 | 6.4 |
| [FLM][RK][RK]p[ST] | NIMA | 13 | 8.6 | 2.26 |

Appendix 3 A Motif extraction for serine phosphorylation (Hepa1-6 phosphoproteome from project 1) by Motif-X algorithm.

| Motif | Motif Score | Foreground Matches | Foreground Size | Fold Increase |
|-----------------------------|-------------|--------------------|-----------------|---------------|
| R-X-X- pS -P-X-P | 37.77 | 40 | 3653 | 20.02 |
| pS -P-X-K | 31.95 | 116 | 3613 | 10.27 |
| K-X-X-X-X-P-X- pS -P | 37.35 | 22 | 3497 | 23.48 |
| P-X- pS -P | 27.79 | 182 | 3475 | 7.14 |
| R-X-X- pS -P | 27.51 | 98 | 3293 | 8.63 |
| pS -P-X-R | 26.40 | 85 | 3195 | 8.39 |
| pS -D-X-E-X-E | 41.56 | 67 | 3110 | 37.98 |
| pS -P-X-X-X-K | 24.58 | 65 | 3043 | 8.44 |
| R-X-R-X-X- pS -X-S | 36.55 | 41 | 2978 | 14.04 |
| D- pS -E-X-E | 44.40 | 49 | 2937 | 29.84 |
| K-X-X-X-X- pS -P | 23.51 | 55 | 2888 | 8.41 |
| R-R-X- pS | 32.00 | 115 | 2833 | 8.96 |
| D-X- pS -D-X-E | 38.53 | 35 | 2718 | 40.63 |
| R-X-X-X- pS -P | 20.95 | 51 | 2683 | 7.17 |
| pS -D-X-E-D | 39.39 | 32 | 2632 | 40.83 |
| E- pS -E-X-E | 39.08 | 33 | 2600 | 22.81 |
| G-X-X- pS -P | 20.90 | 62 | 2567 | 6.56 |
| R-S-X- pS -X-X-X-L | 35.09 | 25 | 2505 | 17.89 |
| G-X-X- pS -P | 20.90 | 62 | 2567 | 6.56 |
| pS -D-X-E | 32.00 | 98 | 2480 | 12.39 |
| L-X-R-S-X- pS | 29.68 | 23 | 2382 | 16.59 |
| K-X-X- pS -P | 20.44 | 38 | 2359 | 7.67 |
| R-X-R-X-X- pS | 25.76 | 67 | 2321 | 7.76 |
| R-X-X- pS -X-S | 32.00 | 79 | 2254 | 6.63 |
| pS -P | 16.00 | 362 | 2175 | 3.85 |
| D- pS -D-X-D | 43.13 | 38 | 1813 | 50.56 |
| E-X-X- pS -E-X-E | 37.47 | 23 | 1775 | 26.10 |
| pS -X-E-D | 32.00 | 74 | 1752 | 9.30 |
| pS -X-D-E | 32.00 | 53 | 1678 | 10.71 |
| R-X-X- pS -X-D | 22.75 | 32 | 1625 | 9.44 |
| pS -X-X-E-X-E | 26.65 | 51 | 1593 | 6.12 |
| pS -X-X-S-X-X-E | 28.09 | 54 | 1542 | 4.92 |
| R-K-X- pS | 23.75 | 36 | 1488 | 9.05 |
| pS -D-X-D | 28.72 | 43 | 1452 | 10.75 |
| pS -X-E-E | 26.85 | 49 | 1409 | 5.75 |
| K-R-X- pS | 32.00 | 45 | 1360 | 10.42 |
| R-X-X- pS -X-P | 21.63 | 32 | 1315 | 7.53 |
| pS -E-X-D | 24.12 | 32 | 1283 | 8.30 |
| R-R- pS | 25.61 | 33 | 1251 | 7.64 |
| R-X-X- pS | 16.00 | 125 | 1218 | 2.87 |
| D- pS -X-D | 25.39 | 27 | 1093 | 9.95 |
| R-X-X-S-X-X- pS | 27.19 | 43 | 1066 | 5.80 |
| R- pS -X-S | 24.92 | 35 | 1023 | 5.93 |
| pS -X-X-X-E-D | 21.84 | 29 | 988 | 7.05 |
| pS -X-X-E-D | 21.68 | 25 | 959 | 8.18 |
| D-S-X- pS | 18.25 | 24 | 934 | 5.09 |
| pS -L-D | 15.89 | 20 | 910 | 4.93 |
| pS -X-S-D | 16.18 | 24 | 890 | 6.22 |
| G- pS | 10.78 | 118 | 866 | 1.91 |
| pS -S-P | 15.20 | 32 | 748 | 4.49 |
| K- pS | 9.72 | 77 | 716 | 2.19 |
| S-X-X- pS | 8.46 | 119 | 639 | 1.72 |
| pS -X-X-X-X-D | 9.63 | 58 | 520 | 2.51 |
| pS -X-D | 6.80 | 42 | 462 | 2.44 |

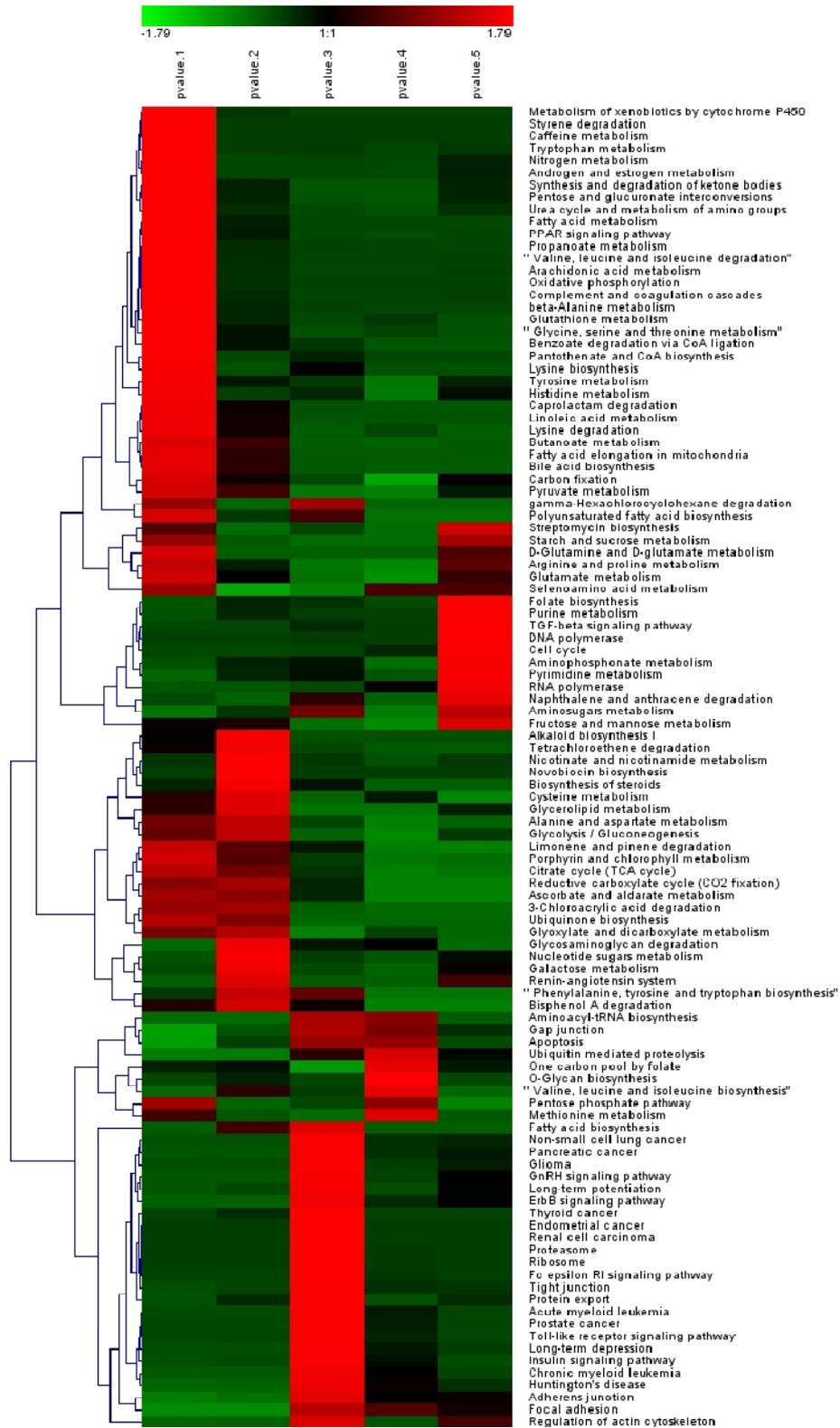
Appendix

| | | | | |
|----------------------|------|----|-----|------|
| E- pS | 7.32 | 53 | 420 | 2.24 |
| pS -L | 5.34 | 71 | 367 | 1.71 |
| S-X- pS | 4.64 | 52 | 296 | 1.80 |
| pS -X-X-X-X-P | 4.93 | 35 | 244 | 2.19 |
| pS -X-S | 4.44 | 39 | 209 | 1.96 |
| R- pS | 4.87 | 24 | 170 | 2.65 |
| R-X- pS | 4.18 | 20 | 146 | 2.66 |

Appendix 3 B Motif extraction for threonine phosphorylation (Hepa1-6 phosphoproteome from project 1) by Motif-X algorithm.

| Motif | Motif Score | Foreground Matches | Foreground Size | Fold Increase |
|-------------------|-------------|--------------------|-----------------|---------------|
| pT -P-P | 22.24 | 33 | 427 | 14.18 |
| P-X- pT -P | 20.67 | 29 | 394 | 11.72 |
| pT -P | 16.00 | 95 | 365 | 4.75 |
| pT -X-X-E | 10.70 | 50 | 270 | 2.86 |
| pT -X-P | 8.03 | 38 | 220 | 2.80 |
| pT -X-X-D | 9.66 | 33 | 182 | 3.57 |
| R-X-X- pT | 5.65 | 24 | 149 | 2.93 |
| S-X- pT | 4.91 | 27 | 125 | 2.43 |

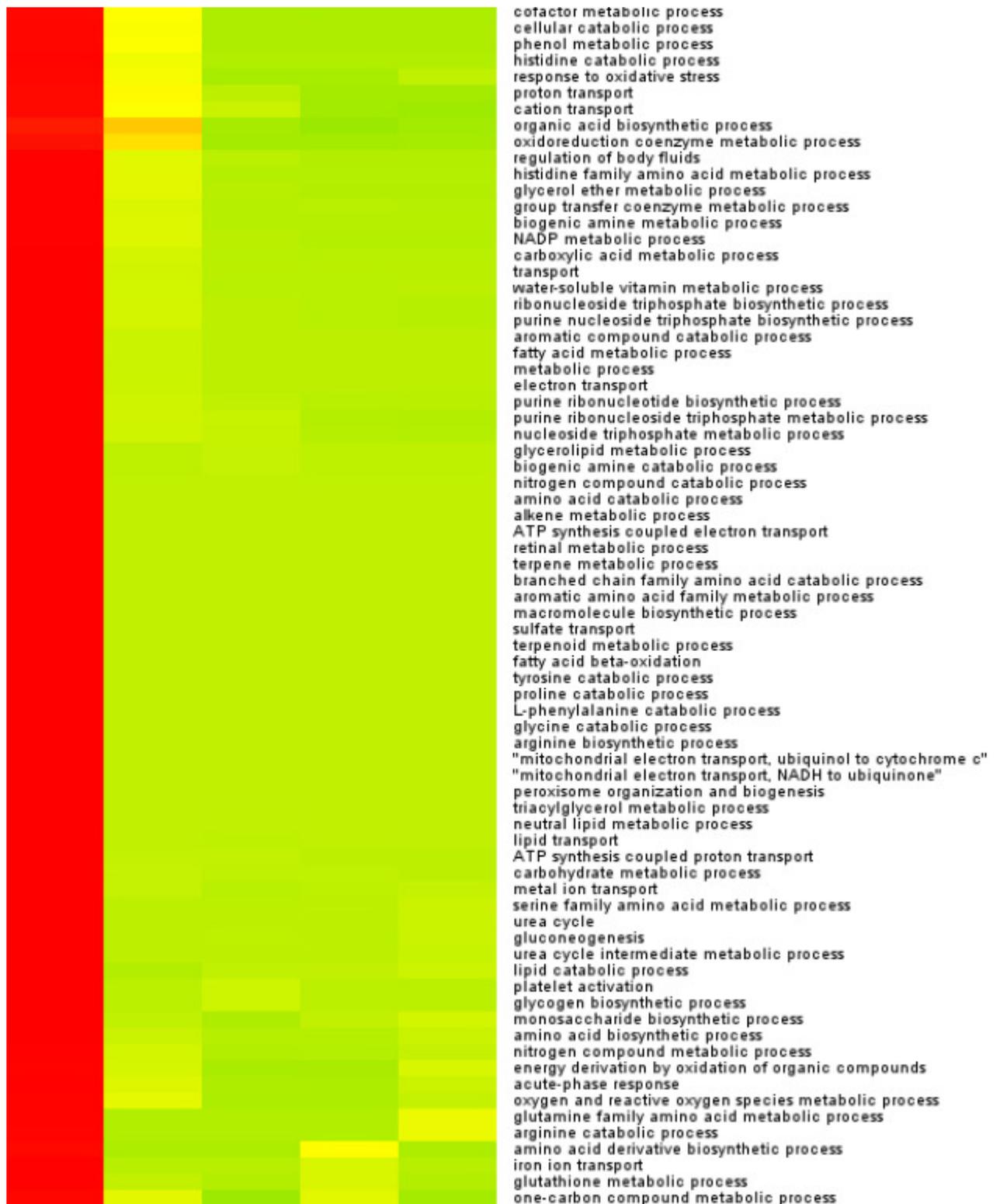
Appendix 4 Primary hepatocyte vs. Hepa1-6 cell line **Sample 1** enriched KEGG pathway categories clustered. Five quantiles according to Figure 4.3.3 were accessed.

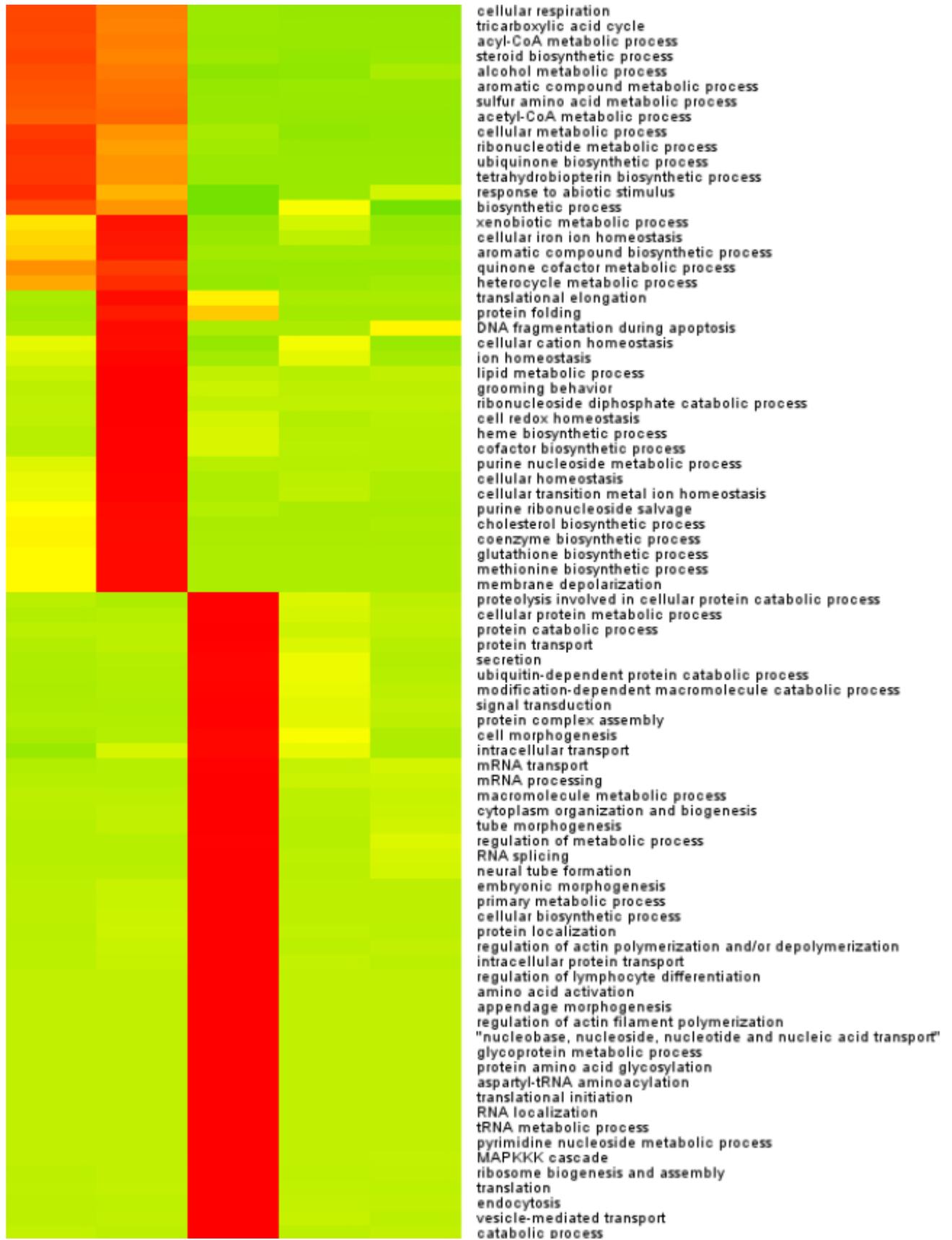


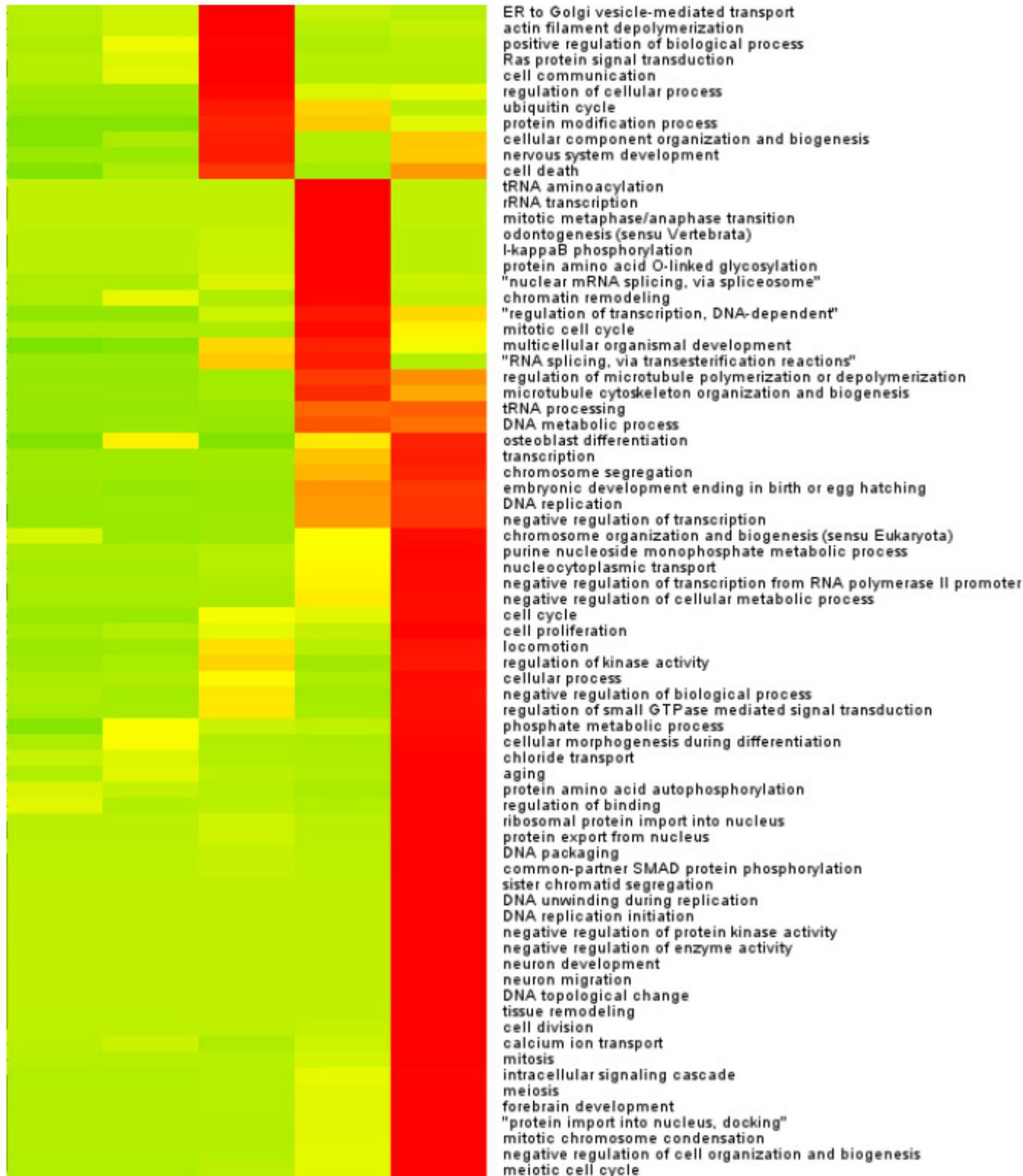
Appendix 5 Primary hepatocyte vs. Hepa1-6 cell line **Sample 1** enriched GO cellular component categories clustered. Five quantiles according to Figure 4.3.3 were accessed.



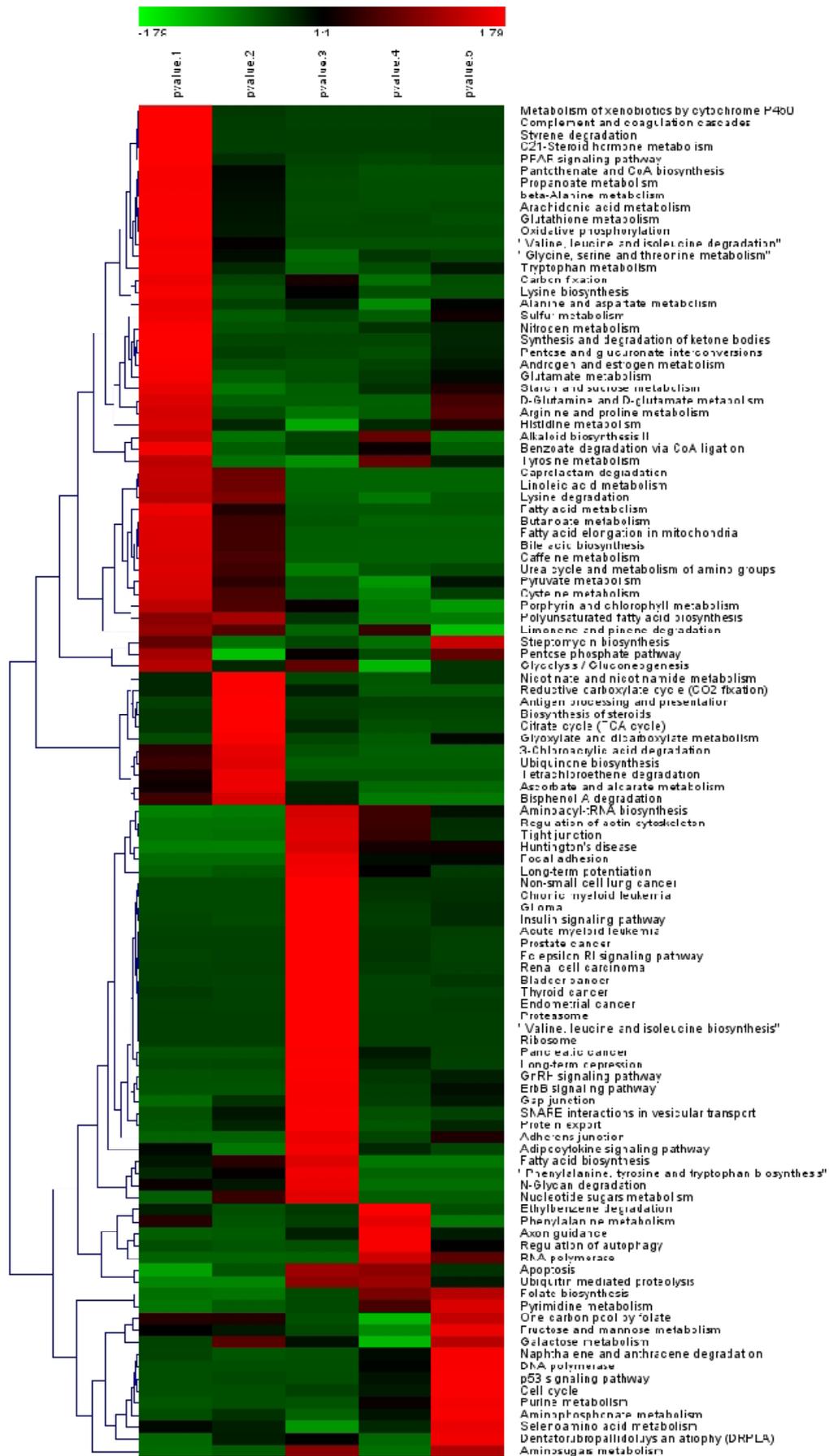
Enlarged Appendix 6







Appendix 7 Primary hepatocyte vs. Hepa1-6 cell line **Sample 2** enriched KEGG pathway categories clustered. Five quantiles according to Figure 4.3.3 were accessed.



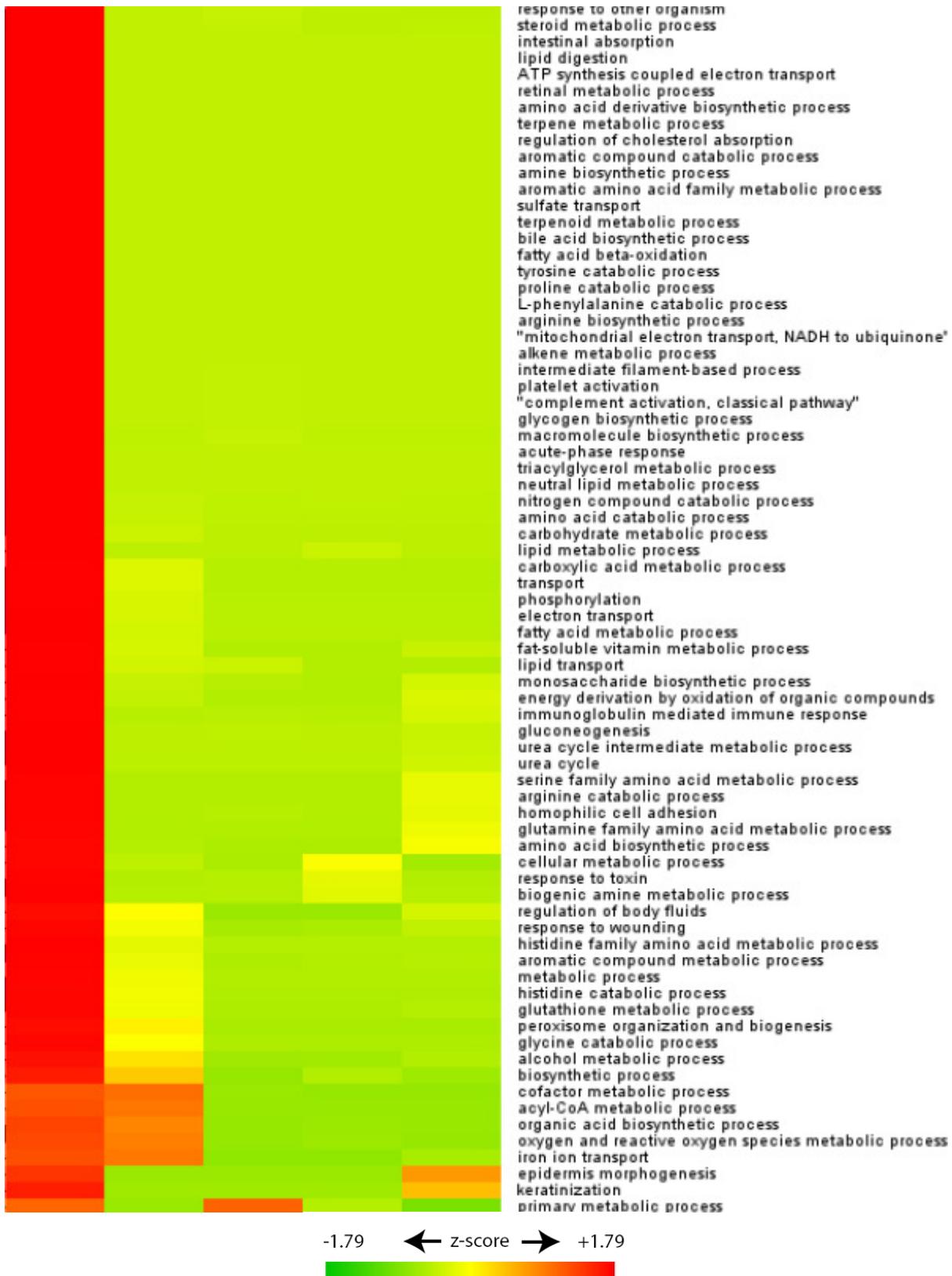
Appendix 8 Primary hepatocyte vs. Hepa1-6 cell line **Sample 2** enriched GO cellular component categories clustered. Five quantiles according to Figure 4.3.3 were accessed.



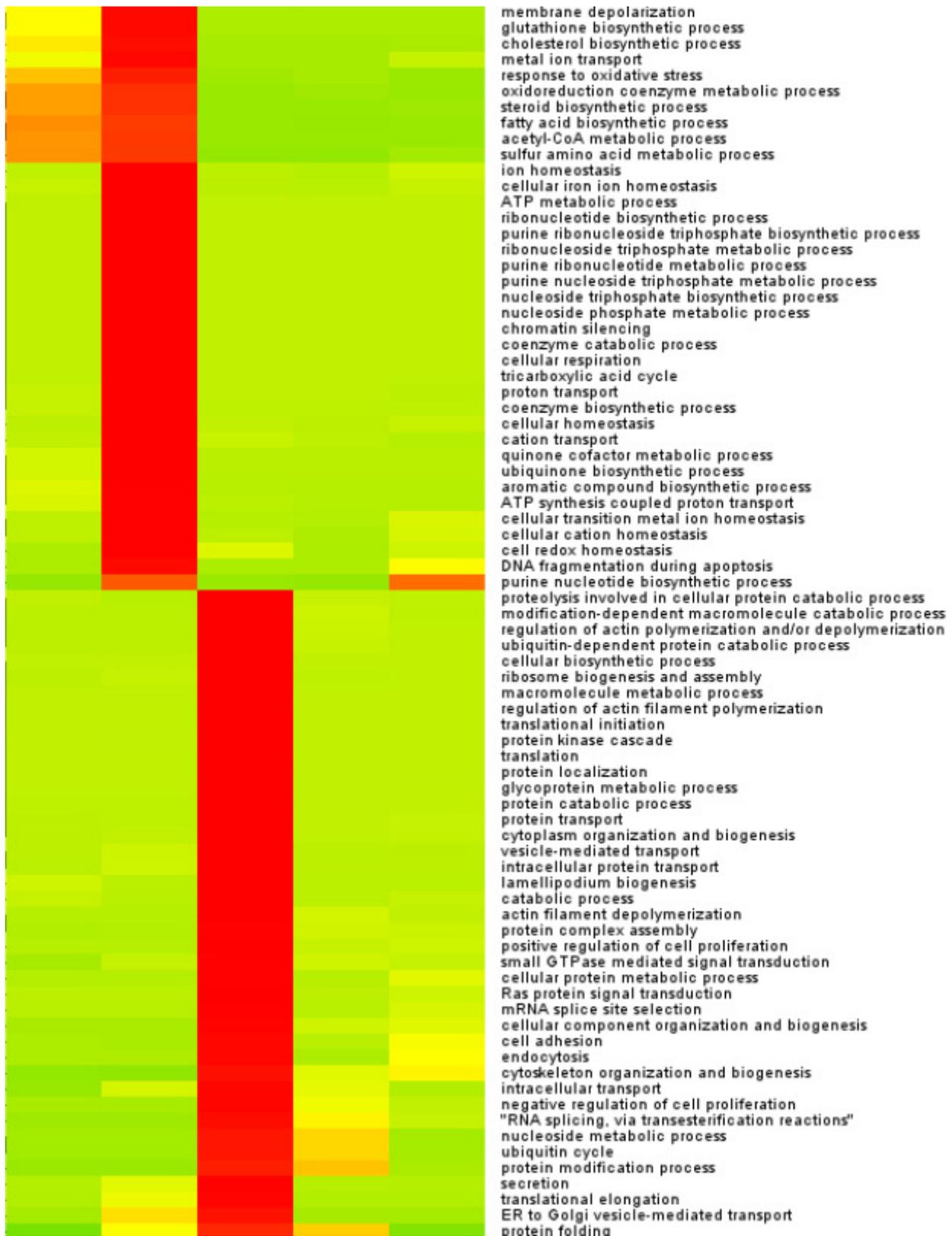
Appendix 9 Primary hepatocyte vs. Hepal-6 cell line **Sample 2** enriched GO biological processes categories clustered. Five quantiles according to Figure 4.3.3 were accessed.



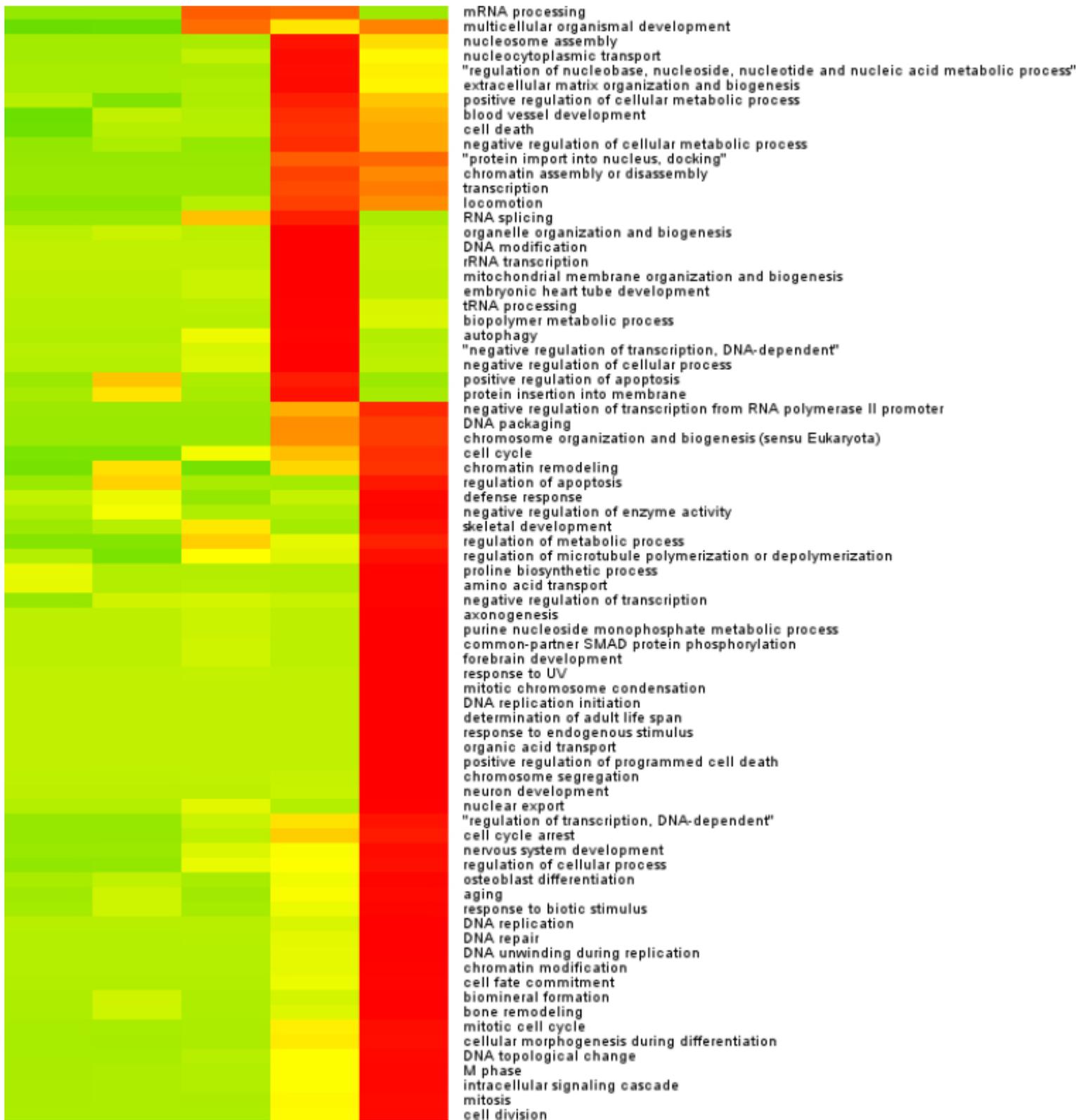
Enlarged Appendix 9



Appendix



Appendix



-1.79 ← z-score → +1.79



Appendix 10 Protein ratios for drug metabolizing enzymes in proteome comparison of primary hepatocyte vs. Hepa1-6 cell line (project 3).

| Family | Protein Names | Uniprot ID | Number of peptides | Number of unique peptides | Ratio H/L Normalized | Ratio H/L Count |
|------------------|---------------|------------|--------------------|---------------------------|----------------------|-----------------|
| Cytochrome P450s | CYP 1A2 | P00186 | 9 | 5 | 0.07 | 3 |
| | CYP 27 | Q9DBG1 | 13 | 13 | 0.02 | 23 |
| | CYP 2A12 | P56593 | 24 | 21 | 0.04 | 29 |
| | CYP 2A4 | P15392 | 5 | 1 | 0.07 | 4 |
| | CYP 2B19 | O55071 | 6 | 1 | 0.01 | 3 |
| | CYP 2B20 | Q62397 | 16 | 11 | 0.09 | 17 |
| | CYP 2C29 | Q64458 | 11 | 6 | 0.04 | 12 |
| | CYP 2C37 | P56654 | 9 | 1 | 0.05 | 4 |
| | CYP 2C40 | P56657 | 6 | 6 | 0.03 | 10 |
| | CYP 2C44 | Q3UEM4 | 3 | 3 | 0.14 | 2 |
| | CYP 2C54 | Q6XVG2 | 9 | 2 | 0.01 | 1 |
| | CYP 2C70 | Q91W64 | 17 | 17 | 0.07 | 9 |
| | CYP 2D10 | P24456 | 12 | 4 | 0.02 | 16 |
| | CYP 2D26 | Q8CIM7 | 16 | 12 | 0.02 | 22 |
| | CYP 2D9 | P11714 | 10 | 6 | 0.02 | 7 |
| | CYP 2E1 | Q05421 | 10 | 10 | 0.03 | 7 |
| | CYP 2F2 | P33267 | 22 | 22 | 0.03 | 31 |
| | CYP 2J5 | O54749 | 3 | 3 | 0.03 | 3 |
| | CYP 3A1 | Q9JKJ9 | 3 | 3 | 0.07 | 3 |
| | CYP 3A11 | Q64459 | 4 | 4 | 0.04 | 3 |
| CYP 3A13 | Q64464 | 6 | 6 | 0.06 | 12 | |
| CYP 4A12 | Q91WL5 | 11 | 11 | 0.07 | 8 | |
| CYP 4F13 | Q99KY6 | 2 | 2 | 0.18 | 2 | |

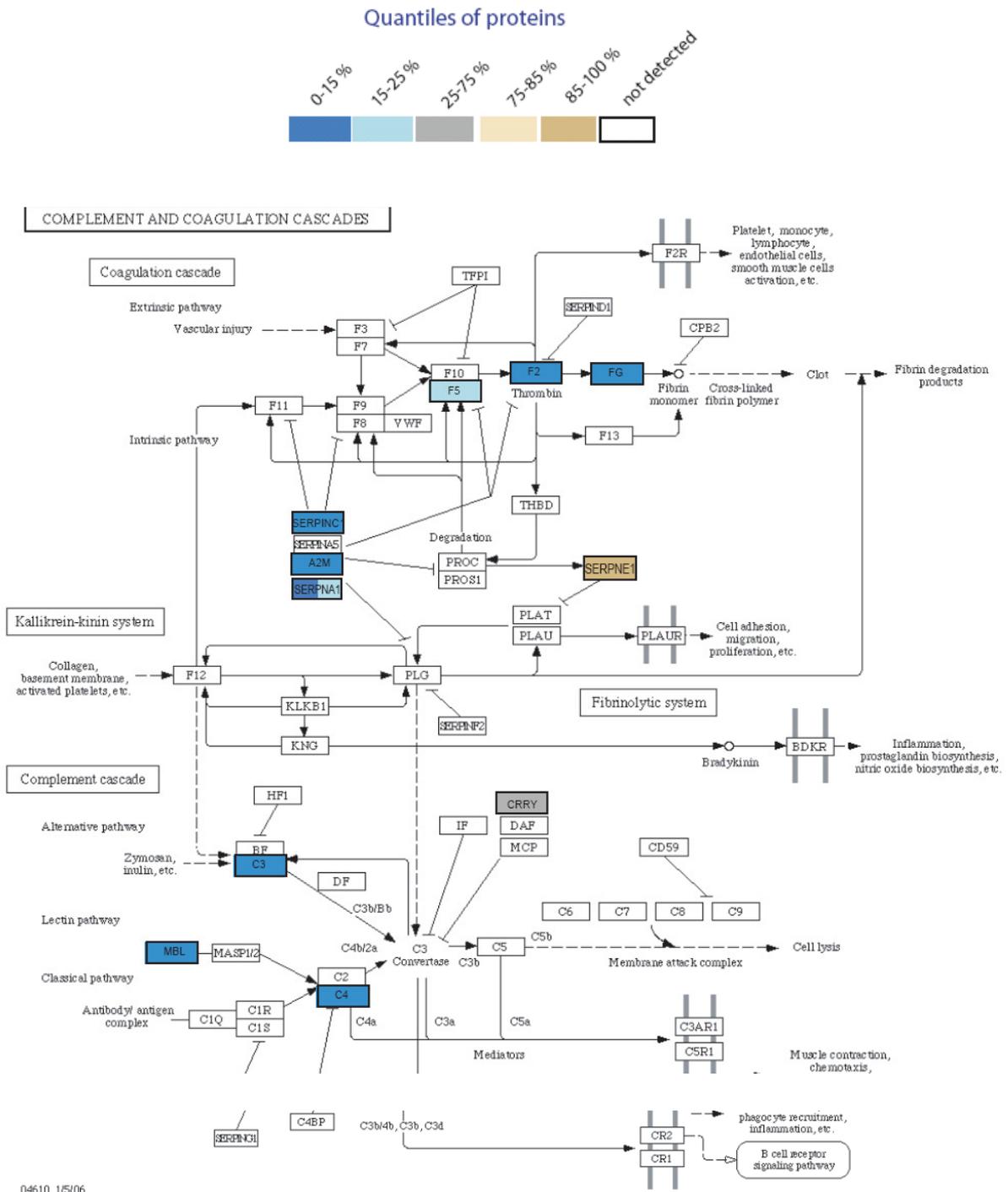
Appendix

| | | | | | | |
|-----------------------------|------------------------------|----------------|----|----|-------|----|
| | CYP 4F14 | Q9EP75 | 2 | 2 | 0.02 | 2 |
| | CYP 4V3 | Q9DBW0 | 6 | 6 | 0.03 | 7 |
| | CYP 7B1 | Q60991 | 10 | 10 | 0.07 | 15 |
| | CYP, 2D22 | Q91W87 | 9 | 5 | 0.02 | 5 |
| | CYP, family 51 | Q8BSQ7 | 4 | 4 | 0.11 | 3 |
| | NADPH--CYP reductase | P37040 | 17 | 17 | 0.44 | 50 |
| | CYP 2S1 | Q9DBX6 | 7 | 7 | 6.54 | 12 |
| | CYP 1A1 | P00184 | 10 | 6 | 0.79 | 21 |
| | RIKEN clone: E130013F06 | Q80Y48 | 4 | 4 | 1.92 | 8 |
| Flavin monooxygenase | Flavin containing | P97872 | 8 | 8 | 0.028 | 10 |
| UDP-glucuronosyltransferase | UGT 1-9 | Q62452 | 6 | 3 | 0.028 | 5 |
| | UGT 1-6 | Q64435 | 12 | 3 | 0.02 | 16 |
| | UGT 1-1 | Q63886 | 10 | 7 | 0.031 | 16 |
| | UGT 2B5 | P17717 | 11 | 2 | 0.067 | 15 |
| | RIKEN clone:C730026L17, | Q3UEI8 | 8 | 5 | 0.01 | 2 |
| | RIKEN clone:C730009A09, | Q3UEC9 | 12 | 8 | 0.014 | 14 |
| Sulfotransferase | Arylsulfotransferase ST1A4 | P52840 | 10 | 10 | 0.019 | 24 |
| | Estrogen sulfotransferase, | P49891 | 2 | 2 | 0.04 | 1 |
| | Sulfotransferase family | Q9QWG7-1 | 4 | 4 | 0.167 | 2 |
| | Tyrosine-ester | O35401 | 7 | 7 | 0.024 | 12 |
| | RIKEN clone:A530080C09 | Q8BGL3 | 11 | 11 | 0.017 | 19 |
| | Heparan sulfate 2-O- | Q8R3H7 | 1 | 1 | 1.788 | 1 |
| glutathione S-transferase | GST A2, A1, alpha 1, alpha 2 | P10648;P13745; | 6 | 3 | 0.034 | 4 |
| | GST alpha 1, alpha 2 | Q6P8Q0;Q6P8Q1; | 6 | 3 | 0.034 | 4 |
| | GST A3 | P30115 | 10 | 7 | 0.008 | 50 |

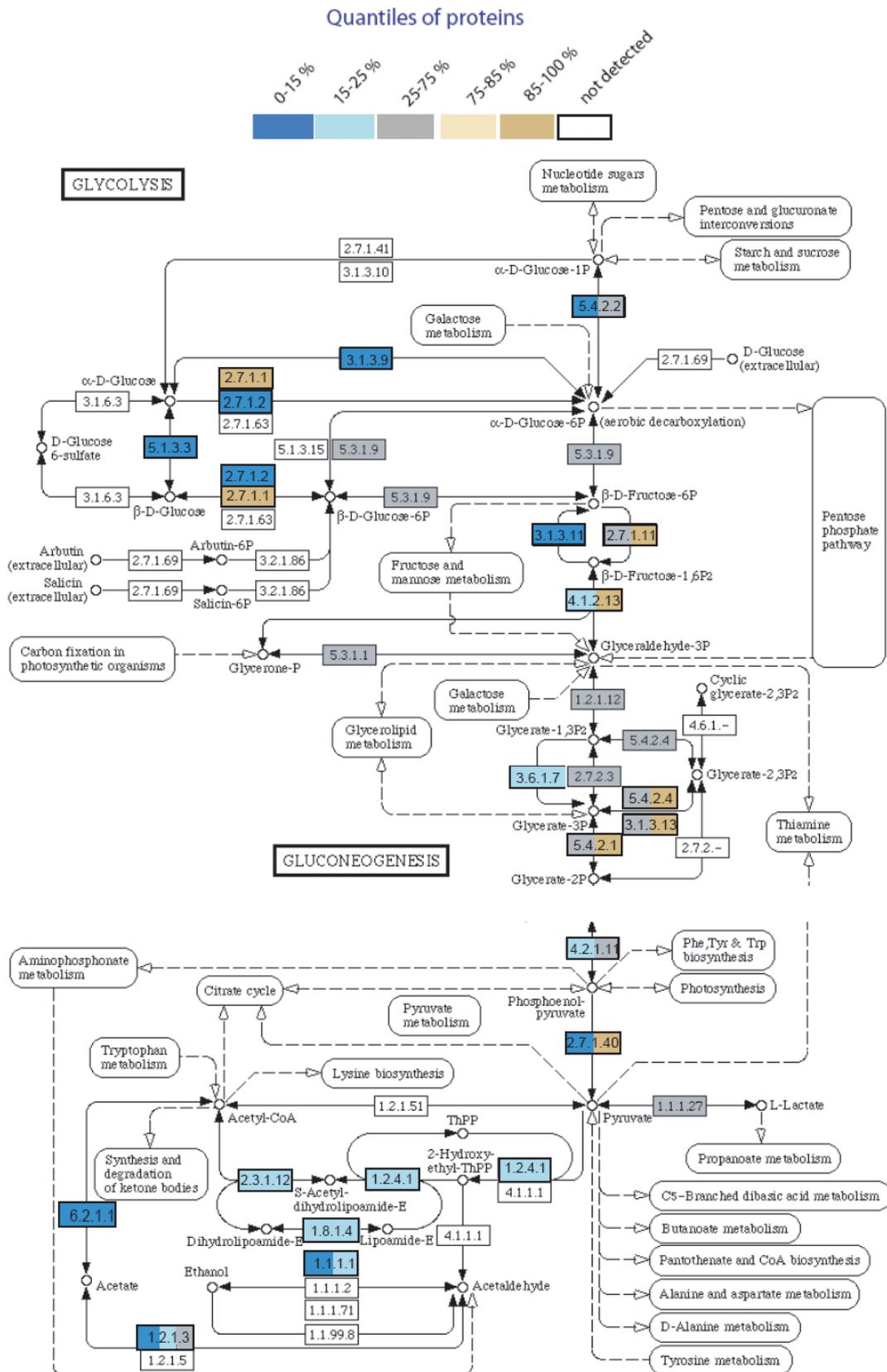
Appendix

| | | | | | | |
|--|------------------|---------------|----|----|-------|-----|
| | GST A4 | P24472 | 5 | 4 | 0.025 | 8 |
| | GST kappa 1 | Q9DCM2 | 14 | 14 | 0.016 | 29 |
| | GST Mu 1 | P10649 | 20 | 13 | 0.046 | 134 |
| | GST Mu 2 | P15626 | 15 | 8 | 0.47 | 40 |
| | GST Mu 3 | P19639 | 12 | 5 | 0.054 | 9 |
| | GST P 1;GST P 2; | P19157;P46425 | 15 | 14 | 0.033 | 196 |
| | GST A3 | P30115 | 10 | 7 | 0.008 | 50 |
| | GST A4 | P24472 | 5 | 4 | 0.025 | 8 |
| | GST kappa 1 | Q9DCM2 | 14 | 14 | 0.016 | 29 |
| | GST Mu 1 | P10649 | 20 | 13 | 0.046 | 134 |
| | GST Mu 2 | P15626 | 15 | 8 | 0.47 | 40 |
| | GST Mu 3 | P19639 | 12 | 5 | 0.054 | 9 |

Appendix 11 KEGG pathway mapping of "complement and coagulation pathways" for proteome comparison of primary hepatocyte vs. Hepa1-6 cell line (project 3).



Appendix 12 KEGG pathway mapping of "glycolysis" for proteome comparison of primary hepatocyte vs. Hepa1-6 cell line (project 3).



Acknowledgements

First of all, I would like to express my deepest gratitude to **Prof. Dr. Matthias Mann** for giving me the opportunity to pursue both of my M.Sc. and Ph.D. studies in his laboratory. Mass spectrometry is an interesting and advanced technology for biology research which has become increasingly important. I am grateful that I had the chance to learn it. Thanks to Matthias for his generous support and supervision. Thanks for thinking for me and working wholeheartedly on my manuscripts for publication among his busiest time. He is an extremely nice person in heart and provides much freedom and many chances to his students for pursuing their scientific interests. His curiousness, diligence and modesty have set a good example and will continue to impose a good influence on me in the future.

Thanks to my coauthors, especially **Chanchal Kumar, Jesper Olsen, Florian Gnad** and **Choudhary Chunaram**. Altogether, your expertise in biology, mass spectrometry and bioinformatics has widened my horizon, given me inspirations and encouraged me to move forward. The feeling of working with you all is great.

I am grateful to **PD. Dr. Ursula Klingmüller** and her students, especially **Sebastian Bohl** in the DKFZ (German Cancer Research Institute), Heidelberg. They are very kind and always available for questions and discussions. The collaboration with them is a sheer joy.

I would like to take the opportunity to thank **Francesca Forner, Boris Macek** and **Peter Bandilla** for getting me started in the practical part of mass spectrometry. Thanks to **Johannes Graumann** and **Jürgen Cox** for sharing their knowledge in bioinformatics and statistics. Johannes' series lectures of R programming are very inspiring. Special thanks to **Yong Zhang**, who provided small scripts to aid my data analysis before the big software MaxQuant was established. Thanks to **Sonja Krüger**, who is always so kind and helpful. Thanks for taking care of the cells while I was away from the lab.

Thanks to many of my lab-mates, especially **Leonie Waanders**, **Lyris de Godoy**, **Michael Lund Nielsen**, **Tiziana Bonaldi**, **Michiel Vermeulen** and **Boris Macek**, for their encouragements during those difficult periods.

Thanks to my fellow student **Aiping Lu**, for being with me and the LTQ-FT 1 in my first year Ph.D. study to stand those cold winter nights and the pressures that the expensive machine didn't work properly... It was really great to have you there. Thanks to other fellow students, **Na Sun**, **Leonie Waanders**, **Ivan Matic**, **Nagarjuna Nagaraj**, **Stefan Hanke**, **Christian Lubner**, **Chanchal Kumar**, **Roland Knispel**, and **Hao-Ven Wang** and others for your friendships! The strengths each of you displayed are impetus to keep me progressing. There are more friendships to remember. To **Tine Klitmøller**, **Sara Zanivan**, **Tanja Bange**, **Gabriele Stoehr** and others, thanks for those nice times together.

Being one of the first batch students of IMPRS graduate school, I am very grateful to the coordinator **Dr. Hans-Jörg Schaffer** who fosters his IMPRS students with enthusiasm and patience. He is inviting and always available to share success and failures. Thank you for your encouragements, Hans-Jörg!

Language mistakes in this thesis were corrected by **Prof. Matthias Mann** and **Dr. Charo Robles**. Among their busiest time of work, they checked my thesis draft sentence by sentence. They also gave me many suggestions in the scientific contents, by pointing out the aspects that I had ignored in the first instance and those factors that I was not aware of. The discussions with them in a broad context have been very beneficial. Here I express my deep gratitude to them. **Dr. Leonard Foster** has given me general suggestions and I thank him for his efforts and kindness.

At last, to my beloved **parents**, **younger sister Cuiting** and boyfriend **Yizi Wu** in China, thousands of thanks for your deep understanding and constant supports! Without these, I would not have achieved so much.

Curriculum Vitae

Personal Information

Name: Cuiping Pan

Gender: female

Nationality: P. R. China

Date of birth: 1981/01/09

Place of birth: GuangDong Province, P. R. China



Academic Qualification

Education:

1999-2003 undergraduate study in Biotechnology, College of Life Sciences, Zhejiang University, Hangzhou, P. R. China

2003- 2005 Master of Science study in Biotechnology, Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark

2005- 2008 Ph.D. study, Department of Proteomic and Signal Transduction, Max-Planck Institute for Biochemistry; International Max-Planck Research School (MPRS) for Molecular and Cellular Life Sciences; Munich, Germany

Research Experiences:

2002-2003 Research project '*Numerical Analysis and Simulation of EEG brain wave*' for 'National Challenge Cup Competition for undergraduates and graduates in Science and Technology'. Supervisors: Changjiang (distinguished) Prof. Bo Zheng and Academician Xiaowei Tang.

2003 Project for B.Sc.: "*Differentially expressed proteins in colorectal cancer identified by HPLC and SELDI-TOF Mass Spectrometry*" in the Cancer Research Institution of Zhejiang University. Supervisor: Changjiang (distinguished) Prof. Xun Hu.

2004 -2005 Project for M.Sc.: "*Quantitative Proteomics to Compare the Tyrosine Phosphorylproteomes in Insulin and IGF-1 Signaling Pathways*" in the Center for

Experimental Bioinformatics of University of Southern Denmark. Supervisors: Dr. Irina Kratchmarova and Prof. Matthias Mann.

2005-now Projects for Ph.D.: “*Quantitative phosphoproteome analysis of a liver cell line reveals specificity of phosphatase inhibitors*”, “*Global effects of kinase inhibitors on signaling networks revealed by quantitative phosphoproteomics*”, and ‘*Proteomic phenotyping*’ to assess differences between transformed and non-transformed mouse liver cells”, in the Department of Proteomics and Signal Transduction, Max-Planck Institute of Biochemistry. Supervisor: Prof. Matthias Mann

Publications:

C.P. Pan, B. Zheng, Y.Z. Wu, Y. Wang & X.W. Tang. Dynamics of Human Brain Electroencephalogram and Alzheimer’s Disease. *Phys. Lett. A* 329, 130-135 (2004)

Cuiping Pan, Quantitative Proteomics to Compare the Tyrosine Phosphoproteomes in insulin and IGF-1 signaling pathways. Master’s thesis (2005)

Schröder B, Wrocklage C, **Pan C**, Jäger R, Kösters B, Schäfer H, Elsässer HP, Mann M & Hasilik A. Integral and associated lysosomal membrane proteins. *Traffic*. 8(12):1676-86 (2007)

Cuiping Pan, Florian Gnad, Jesper V. Olsen & Matthias Mann. Quantitative phosphoproteome analysis of a liver cell line reveals specificity of phosphatase inhibitors. *Proteomics* (accepted manuscript)

Awards

Top Honor of ‘Zhejiang Province Challenge Cup Competition for undergraduates and graduates in Science and Technology’ (June, 2003, China)

Young Investigator Awardee of the 3rd Human Proteome Organization (HUPO) Conference (Oct, 2004, China)