

**Development of Mass Spectrometry Based Technologies for Quantitative
Cell Signaling Phosphoproteomics:
The Epidermal Growth Factor Receptor Family as a Model System**

by

Alejandro Wolf Yadlin

B.S Biotechnology Engineering Universidad de Chile, 1999
Ing. Biotechnology Engineering Universidad de Chile, 2001
M.S Chemical Engineering Universidad de Chile, 2001

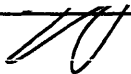
SUBMITTED TO THE DIVISION OF BIOLOGICAL ENGINEERING IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN BIOENGINEERING
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
[February 2007]
SEPTEMBER, 2006

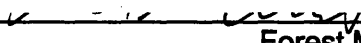
© Alejandro M. Wolf Yadlin. All rights reserved.

The author hereby grants to MIT permission to reproduce
and to distribute publicly paper and electronic
copies of this thesis document in whole or in part
in any medium now known or hereafter created

Signature of Author: _____

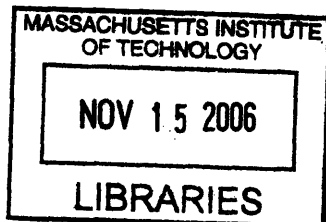

Alejandro M. Wolf Yadlin
Division of Biological Engineering
September 12, 2006

Certified by: _____


Forest M. White
Associate Professor of Biological Engineering
Thesis Supervisor

Accepted by: _____


Douglas A. Lauffenburger
Professor of Biology and Chemical and Biological Engineering
Chair, Biological Engineering Division – Thesis Supervisor



ARCHIVES

**Development of Mass Spectrometry Based Technologies for Quantitative
Cell Signaling Phosphoproteomics:
The Epidermal Growth Factor Receptor Family as a Model System**

by

Alejandro Wolf Yadlin

B.S Biotechnology Engineering Universidad de Chile, 1999
Ing. Biotechnology Engineering Universidad de Chile, 2001
M.S Chemical Engineering Universidad de Chile, 2001

SUBMITTED TO THE DIVISION OF BIOLOGICAL ENGINEERING IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN BIOENGINEERING
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER, 2006

© Alejandro M. Wolf Yadlin. All rights reserved.

The author hereby grants to MIT permission to reproduce
and to distribute publicly paper and electronic
copies of this thesis document in whole or in part
in any medium now known or hereafter created

Signature of Committee Chair: _____
Linda Griffith
Professor of Mechanical and Biological Engineering

Signature of Committee Member: _____
K. Dane Wittrup
Professor of Chemical and Biological Engineering

A la Mamá, el Tata y la Yaya con cariño

ABSTRACT	11
I INTRODUCTION	
I.1 EPIDERMAL GROWTH FACTOR RECEPTOR FAMILY	15
I.1.1 EGFR Family Members Ligands a Brief Description	16
I.1.2 EGFR Family Members Structure and Mechanism of Activation	17
I.1.3 Cell Signaling in the EGFR Family Context	22
I.1.3.1 RAS/MAPK	24
I.1.3.2 PI3K/AKT	26
I.1.3.3 PLC- γ /PKC	29
I.1.3.4 STATs	31
I.1.3.5 Other Pathways	33
I.1.4 Physiological Role of the EGFR Family	37
I.1.4.1 Heart	37
I.1.4.2 Nervous System	37
I.1.4.3 Epithelial development	38
I.1.4.4 Mammary Gland	39
I.1.5 EGFR Family Members Roles in Cancer	40
I.1.5.1 Breast Cancer	41
I.1.5.2 Other Cancers	42
I.1.6 EGFR Family Related Therapeutics Compounds	44
I.2 MASS SPECTROMETRY IN PHOSPHOPROTEOMICS	46
I.2.1 Ion Sources	46
I.2.2 Mass Spectrometers	48
I.2.2.1 Quadrupole Time-of-Flight and IDA	48
I.2.2.2 Triple Quadrupoles and MRM	53
I.2.2.3 QSTAR v/s QTRAP: Resolution v/s Sensitivity and Dynamic Range	57
I.2.3 Proteomics and Phosphoproteomics	59
I.2.3.1 Liquid chromatography coupled mass spectrometry	59
I.2.3.2 Sample Preparation for Phosphoproteomics: Before Mass Spectrometry	60
I.2.3.3 Detection Strategies for Phosphoproteomics	63
I.2.3.4 Quantification Methods for Mass Spectrometry	64
I.3 MOTIVATION	68
I.4 REFERENCES	70

II ANTIBODIES AGAINST DOMAINS II AND IV OF EGFR INHIBIT RECEPTOR PHOSPHORYLATION AND DECREASE HIGH-AFFINITY BINDING BY BLOCKING PREFORMED DIMERS

II.1	SUMMARY	105
II.2	MATERIALS AND METHODS	106
II.2.1	Peptides and antibodies	106
II.2.2	Cell Culture	106
II.2.3	Cell Treatment	107
II.2.4	Protein Quantification	108
II.2.5	Quantitative Western Blot Analysis	108
II.2.6	Flow Cytometry	109
II.3	RESULTS	110
II.3.1	Rabbit polyclonal antisera generation	110
II.3.2	Effect of antisera on receptor activation	111
II.3.3	Effect of antisera on EGF binding	115
II.4	DISCUSSION	117
II.5	REFERENCES	120

III TIME-RESOLVED MASS SPECTROMETRY OF TYROSINE PHOSPHORYLATION SITES IN THE EGF RECEPTOR SIGNALING NETWORK REVEALS DYNAMIC MODULES

III.1	SUMMARY	125
III.2	MATERIALS AND METHODS	127
III.2.1	Cell Culture, EGF stimulation	127
III.2.2	Protein Quantification	127
III.2.3	Cell Lysis, Protein Digestion and Peptide Fractionation	128
III.2.4	iTRAQ Labeling and peptide Immunoprecipitation	128
III.2.5	IMAC and Mass Spectrometry Analysis	129
III.2.6	Western Blot Analysis	130
III.2.7	Phosphopeptide Sequencing, Data Clustering and Analysis	131
III.3	RESULTS	133
III.3.1	Time Resolved Tyrosine Phosphorylation after EGF Stimulation Measured by Quantitative Mass Spectrometry	133
III.3.1.1	Methodology	133
III.3.1.2	Protein Phosphorylation Insights	137
III.3.1.3	Validation	141
III.3.1.4	Data Variability	143
III.3.2	Bioinformatic Analysis Reveals Dynamic Modules within the EGFR Signaling Network	144

III.4	DISCUSSION	151
III.5	REFERENCES	156
IV QUANTITATIVE PHOSPHOPROTEOMIC ANALYSIS OF HER2-OVEREXPRESSION EFFECTS ON CELL SIGNALING NETWORKS GOVERNING PROLIFERATION AND MIGRATION		
IV.1	SUMMARY	163
IV.2	Materials and Methods	165
IV.2.1	Cell Culture and Stimulation	165
IV.2.2	Protein Quantification	166
IV.2.3	Cell Lysis, Protein Digestion, Peptide Fractionation and iTRAQ Labeling	166
IV.2.4	Peptide Immunoprecipitation	167
IV.2.5	IMAC and Mass Spectrometry	168
IV.2.6	Phosphopeptide Sequencing, Data Clustering and Analysis	168
IV.2.7	ELISA protocol for ErbB3 receptor quantification	169
IV.2.8	Hierarchical Clustering	170
IV.2.9	Self Organizing Maps	170
IV.2.10	Proliferation	172
IV.2.11	Migration	173
IV.2.12	Partial Least Squares Regression	174
IV.3	RESULTS	176
IV.3.1	Phosphotyrosine mass spectrometry of parental and HER2-overexpressing HMECs in response to EGF and HRG treatment	176
IV.3.2	Hierarchical Clustering reveals tyrosine phosphorylation sites co-regulated across conditional space	180
IV.3.3	Self-Organizing Maps define temporal and conditionally related clusters of phosphorylation sites	184
IV.3.4	Cell proliferation and migration are differentially stimulated via EGFR and HER2	190
IV.3.5	Modulation of EGFR signaling network by HER2	192
IV.3.6	HRG vs EGF stimulation in the presence of HER2	198
IV.3.7	PLSR modeling correlates signals with cell functional responses	202
IV.4	DISCUSSION	207
IV.5	REFERNCES	210

V MULTIPLE REACTION MONITORING AND INFORMATION DEPENDENT ACQUISITION: TOWARDS A REPRODUCIBLE, QUANTITATIVE AND ROBUST MASS SPECTROMETRY BASED TECHNOLOGY FOR PROTEOMICS

V.1	SUMMARY	221
V.2	MATERIALS AND METHODS	224
V.2.1	Cell Culture and Stimulation	224
V.2.2	Protein Quantification	224
V.2.3	Sample Preparation	225
V.2.4	Peptide IP	225
V.2.5	IMAC and IDA Mass Spectrometry Analysis	226
V.2.6	MRM Mass Spectrometry Analysis	227
V.2.7	Phosphopeptide Sequencing, Data Clustering and Analysis	228
V.3	RESULTS	231
V.3.1	A Robust Technology for Mass Spectrometry Based Phosphoproteomics	231
V.3.2	QTRAP/CSDMP Data Output	236
V.3.3	Quantification Quality	239
V.3.4	Reproducibility	243
V.3.5	Why 7 time points?	246
V.4	DISCUSSION	249
V.4.1	Resolution and Sensitivity	249
V.4.2	Parent ion isolation and charge states	251
V.4.3	LC elution and MRM mode limitations	251
V.5	REFERENCES	255
	CONCLUSIONS	257
	ACKNOWLEDGEMENTS	261
	DEDICATIONS	263
	INDEX OF TABLES AND FIGURES	264
	ABBREVIATIONS	269
	PERMISSIONS	271

**Development of Mass Spectrometry Based Technologies for Quantitative Cell Signaling Phosphoproteomics:
The Epidermal Growth Factor Receptor Family as a Model System**

by

Alejandro Wolf Yadlin

Submitted to the division of biological engineering in partial fulfillment of the requirements for the degree of doctor of philosophy in bioengineering at the Massachusetts Institute of Technology

ABSTRACT

Ligand binding to cell surface receptors initiates a cascade of signaling events regulated by dynamic phosphorylation on a multitude of pathway proteins. Quantitative features, including intensity, timing, and duration of phosphorylation of particular residues play a role in determining cellular response. Mass spectrometry has been previously used to identify and catalog phosphorylation sites or quantify the phosphorylation dynamics of proteins in cell signaling networks. However, identification of phosphorylation sites presents little insight on cellular processes and quantification of phosphorylation dynamics of whole proteins masks the different roles that several phosphorylation sites within one protein have in the network. We have designed a mass spectrometry technique allowing site-specific quantification of dynamic phosphorylation in the cell. We have applied this technique to study signaling events triggered by different members of the epidermal growth factor receptor (EGFR) family. Self organizing maps (SOMs) analysis of our data has highlighted potential biological functions for phosphorylation sites previously unrelated to EGFR signaling and identified network modules regulated by different combinations of EGFR family members. Partial least square regression (PLSR) analysis of our data identified combination of signals strongly correlating with cellular proliferation and migration. Because our method was based on information dependent acquisition (IDA) the reproducibility of peptides identified across multiple analyses was low. To improve our methodology to permit both discovery of new phosphorylation sites and robust quantification of hundreds of nodes within a signaling network we combined IDA-analysis with multiple reaction monitoring (MRM) of selected precursor ions. MRM quantification of high resolution temporal profiles of the EGFR network provided 88% reproducibility across four different samples, as compared to 34% reproducibility by IDA only.

In summary, we have developed a new robust mass spectrometry technique allowing site specific identification, quantification and monitoring of dynamic phosphorylation in the cell with high temporal resolution and under any number of biological conditions. Because the data obtained with this method is not sparse it is especially well suited to mathematical and computational analyses. The methodology is also broadly applicable to multiple signaling networks and to a variety of samples, including quantitative analysis of signaling networks in clinical samples.

Thesis Supervisor: Douglas A. Lauffenburger

Title: Professor of Biological Engineering

Thesis Supervisor: Forest M. White

Title: Associate Professor of Biological Engineering

I INTRODUCTION

Brief Review of the EGFR Family Signaling Network and Mass Spectrometry Technologies for Phosphoproteomics

I.1 EPIDERMAL GROWTH FACTOR RECEPTOR FAMILY

Originally named ERBB for its homology to the erythroblastoma viral gene product v-erbB, the epidermal growth factor receptor (EGFR) family of proteins comprises four transmembrane receptor tyrosine kinases (RTK) (EGFR/ERBB1/HER1, ERBB2/HER2/neu, ERBB3/HER3, ERBB4/HER4) and at least 12 ligands, all of them containing an epidermal growth factor (EGF) conserved motif (Citri & Yarden, 2006; Yarden & Sliwkowski, 2001).

Since the discovery and isolation of the epidermal growth factor by S.Cohen (Cohen, 1965; Cohen & Elliott, 1963; Levi-Montalcini & Cohen, 1960), the detection and cloning EGFR (Cohen & Carpenter, 1975; Cohen et al., 1982; Xu et al., 1984), the cloning of HER2 in 1985, HER3 in 1989 and HER4 in 1993 (Kraus et al., 1989; Plowman et al., 1993; Semba et al., 1985) and the discovery of EGFR and HER2 gene amplification and overexpression in human cancer cells during the 1980s (King et al., 1985; Lacroix et al., 1989; Perez et al., 1984; Yokota et al., 1986; Zeillinger et al., 1989) research in the EGFR family field has grown exponentially as scientists try to understand the mechanisms of cell signaling and find a cure for cancer.

In this section I will introduce the EGFR family members ligands and receptors, activation mechanisms, signaling cascades, physiological functions and their role in pathologies, with particular focus on breast cancer.

1.1.1 EGFR Family Members Ligands a Brief Description

The – at least – 12 ligands that bind the EGFR family members can be divided into three groups. The first group includes epidermal growth factor (EGF), transforming growth factor- α , amphiregulin (AR) and epigen and binds exclusively to EGFR (Harris et al., 2003). The second group includes heparin-binding EGF-like growth factor (HB-EGF), betacellulin (BTC), and epiregulin (EPR) and can bind both EGFR and HER4 (and in the case of HB-EGF also HER3) (Beerli & Hynes, 1996; Elenius et al., 1997; Piepkorn et al., 1998). The third and final group binds only HER3 and HER4 and is comprised by the neuregulins (NRGs) (or heregulins (HRG)). This group can be further subdivided in peptides that bind HER3 and HER4: Heregulin (HRG/NRG1) and NRG2 and peptides that bind only HER4: NRG3, NRG4 and tomoregulin (Falls, 2003; Normanno et al., 2006).

All EGFR family ligands are polypeptides produced as type I transmembrane proteins. In order to release mature growth factors that can bind the receptors (although juxtacrine binding by uncleaved ligands is possible too) the pro-growth factors are cleaved by proteases at the cell surface.

EGFR family ligands can be distinguished by a consensus sequence, known as the EGFR motif, in which six spatially conserved cysteine residues form three disulfide bonds within the peptide (C1-C3, C2-C4, C5-C6). The EGF motif – specifically the tertiary structure it generates – is essential for binding EGFR family members (Harris et al., 2003).

1.1.2 EGFR Family Members Structure and Mechanism of Activation

The four EGFR family receptors (EGFR, HER2, HER3 and HER4) share similar architecture. An extracellular domain, characterized by the presence of two rigid, cysteine-rich (domains II and IV) and two flexible, leucine-rich (domains I and III) sub-domains, able to – homo or hetero – dimerize with other family members (Bouyain et al., 2005; Cho & Leahy, 2002; Cho et al., 2003; Ferguson et al., 2003; Garrett et al., 2003; Garrett et al., 2002; Ogiso et al., 2002) an hydrophobic single chain transmembrane domain and a cytosolic section containing a very conserved kinase domain and the receptor tyrosine auto-phosphorylation sites at the C-terminal loop (Stamos et al., 2002).

Out of the four HER receptors only EGFR and HER4 have kinase activity and are capable of ligand binding (Yarden & Sliwkowski, 2001). HER2 is unable to bind any ligand as confirmed by its crystal structure (Garrett et al., 2003; Klapper et al., 1999) and HER3 lacks kinase activity due to amino acid substitution in critical residues of its catalytic domain; specifically C721, H740 and N815 correspond to A, Q and D in all other known protein tyrosine kinases (Citri et al., 2003; Garrett et al., 2003; Guy et al., 1994; Sierke et al., 1997). Despite their “shortcomings” HER2 and HER3 receptors are capable of potent signaling through heterodimer formation with other EGFR family members and notably among themselves (Citri et al., 2003; Citri & Yarden, 2006; Yarden & Sliwkowski, 2001).

Of particular interest for this thesis are the characteristics of the extracellular structure of these receptors. As model for the whole family I will focus on the structure of EGFR (ligand binding) and HER2 (non-ligand binding) ectodomains. For details on the structure of EGFR kinase domain the work by Stamos et al. (Stamos et al., 2002) is recommended. For details on the HER3 and HER4 ectodomain structure the works by Cho and Leahy and Bouyain et al. (Bouyain et al., 2005; Cho & Leahy, 2002) are recommended.

EGFR (and also HER3 and HER4) extracellular domain can be found in two conformations which have been designated as extended/active and tethered/auto-inhibited respectively. In the active conformation domains I and III bind the ligand (Figure I.1) while domain II and IV protrude parallel to the hypothetical cell surface. The crystal structure of EGF bound to EGFR reveals that the signaling complex is formed by 2 receptors and 2 ligands and that dimerization is receptor mediated through EGFR domain II (Burgess et al., 2003; Ogiso et al., 2002). Although domain IV is not resolved in the structure, it might participate in dimer formation since peptides mimicking portions of domain IV can inhibit dimerization (Berezov et al., 2002), nonetheless its role could be only minimal with a recent study showing that more than 90% of the binding energy required for dimer formation is accounted by domain II interactions (Dawson et al., 2005).

The EGFR crystallized monomer (Figure I.2) shows its auto-inhibited state, where domain II is bound to domain IV and unable to mediate dimer formation as in Figure I.1. Interestingly Y246 a key residue for dimerization is

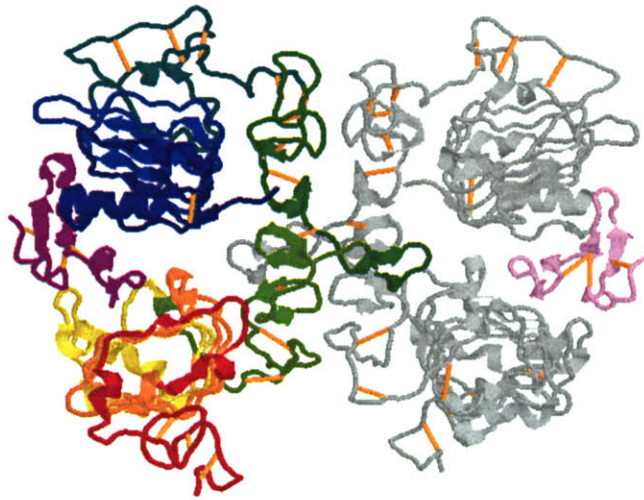


Figure I.1 – Crystal structure of EGFR-EGF homodimer. Blue, domain I. Green domain II. Yellow/red domain III and tail of IV. Domain IV is truncated. Magenta EGF (binding between domains I and III). Orange rods represent di-sulfide bonds.

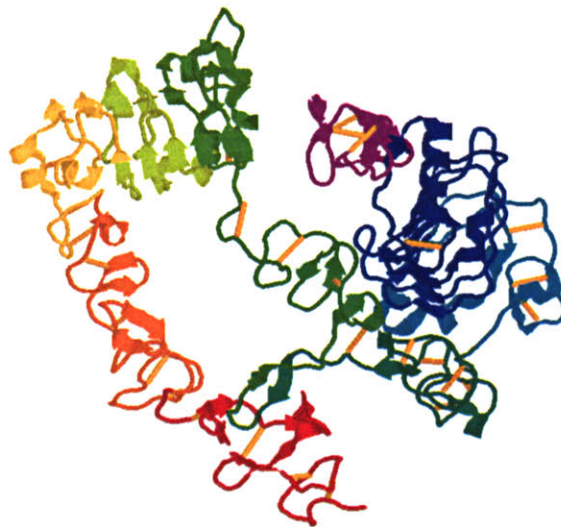


Figure I.2 – Crystal structure of auto-inhibited EGFR bound to EGF. Blue, domain I. Green domain II. Yellow domain III. Red domain IV. Magenta EGF (binding only domain I). Orange rods represent di-sulfide bonds.

also at the heart of the auto-inhibition contacts (Ferguson et al., 2003). The tethered configuration is also characterized by the distance between domains I

and III, which makes impossible their simultaneous binding by EGFR ligands (Ferguson et al., 2003), as occurs in the extended conformation (see Figure I.1).

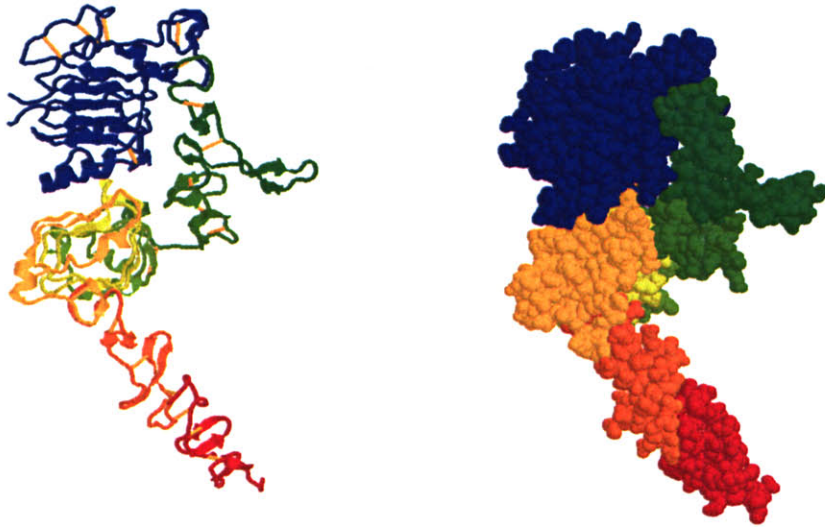


Figure I.3 – Crystal structure of HER2. Blue, domain I. Green domain II. Yellow domain III. Red domain IV. Orange rods represent di-sulfide bonds. Notice from the space fill structure at the right that there is no room for ligand binding between domains I and III.

The prevalent hypothesis in the field is that EGFR exists on the surface of the cell in a thermodynamic equilibrium between the tethered and open states, the first one being highly favored at any given time (more than 90% of the receptors are present in a tethered conformation at any given time). EGF/ligand binding is thought to stabilize the open configuration shifting the equilibrium towards the active state, revealing domain II dimerization arm and allowing dimer formation. Consistent with this hypothesis the crystal structure of HER2 – the most promiscuous receptor of the family (Citri et al., 2003) – shows its inability to adopt the tethered configuration and bind ligand. In fact, the crystal structure of HER2 displays significant contacts between domains I and III, leaving no space

for ligand binding and locking the receptor in the open configuration (Figure 1.3) (Burgess et al., 2003).

Although the crystal structures give great insight into the activation process of EGFR family members, further studies have shown that disruption of the tethered configuration is necessary, but not sufficient to activate EGFR signaling (Mattoon et al., 2004; Walker et al., 2004). It appears that to allow efficient dimerization ligand binding is necessary to orientate the residues C-terminal of domain II dimerization arm (Dawson et al., 2005). Preformed dimers had also been observed in the absence of ligands by fluorescence microscopy (Gadella & Jovin, 1995), single molecule tracking (Sako et al., 2000), fluorescence resonant energy transfer (FRET) (Martin-Fernandez et al., 2002) and cross-linking studies (Yu et al., 2002). In any case these dimers show only basal phosphorylation states (Yu et al., 2002) suggesting that there might be active and inactive dimer configurations, with ligand binding still the key step for activation (Moriki et al., 2001).

Due to the importance of ligand binding for receptor activation many studies have been conducted to measure the kinetic and equilibrium constants of EGF EGFR binding. Cell surface titrations of EGF-EGFR binding reveal two populations of receptors, one with high affinity and other with low affinity for EGF (Lax et al., 1989). While the high affinity receptors ($K_d = 100\text{-}300\text{ pM}$) represent about 5-10% of the total population, the rest of the receptors bind EGF with $K_d = 2\text{-}15\text{ nM}$ (Schlessinger, 1988). Early experiments suggested that the observed high affinity component was due to EGF binding to EGFR dimers (Boni-

Schnetzler & Pilch, 1987; Yarden & Schlessinger, 1987). In fact, models based on the experimental information were able to produce concave-up Scatchard plots qualitatively describing the data, but they were unable to fit the data in a quantitative fashion (Wofsy et al., 1992). Only more refined models taking account heterogeneity of receptor density on the cell surface (Mayawala et al., 2005; Wofsy et al., 1992) could fit the data. Further models have been developed where receptor binding to – a probably not biologically relevant – hypothetical third site has been used to quantitatively fit the experimental observations (Holbrook et al., 2000; Klein et al., 2004). Although the fit of these models is good, it is hard to assess the biological intuition behind the third binding site.

1.1.3 Cell Signaling in the EGFR Family Context

Cell signaling downstream of receptor tyrosine kinases, such as the EGFR family members, comprises an interconnected network of pathways associated with various regulatory processes. These pathways involve a number of components whose interactions are required to carry out regulatory processes and that are themselves regulated by covalent modifications, including phosphorylation on tyrosine as well as serine and threonine residues. As seen above – with the exception of HER2 – in the case of EGFR family members, ligand binding activates the receptor and dimerization results in autophosphorylation of selected tyrosine sites in the C-terminal region (Holbro &

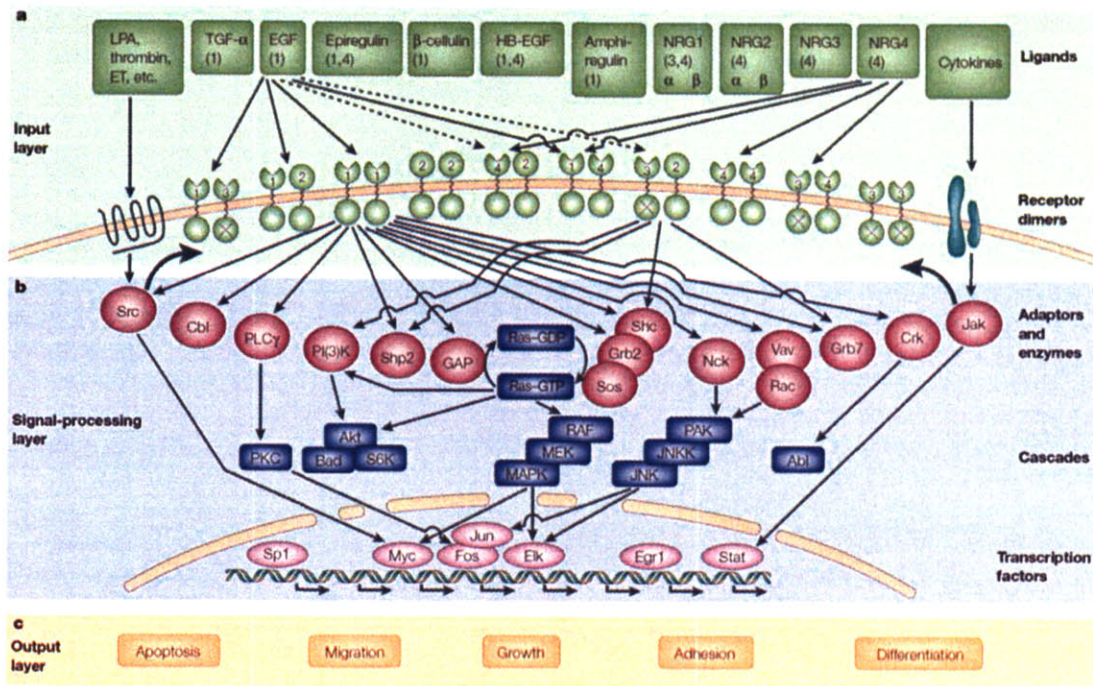


Figure 1.4 – The ErbB signaling network. **a** | Ligands and the ten dimeric receptor combinations comprise the input layer. Numbers in each ligand block indicate the respective high-affinity ErbB receptors. For simplicity, specificities of receptor binding are shown only for epidermal growth factor (EGF) and neuregulin 4 (NRG4). ErbB2 binds no ligand with high affinity, and ErbB3 homodimers are catalytically inactive (crossed kinase domains). *Trans*-regulation by G-protein coupled receptors (such as those for lysophosphatidic acid (LPA), thrombin and endothelin (ET)), and cytokine receptors is shown by wide arrows. **b** | Signaling to the adaptor/enzyme layer is shown only for two receptor dimers: the weakly mitogenic ErbB1 homodimer, and the relatively potent ErbB2–ErbB3 heterodimer. Only some of the pathways and transcription factors are represented in this layer. **c** | How they are translated to specific types of output is poorly understood at present. (Abl, a proto-oncogenic tyrosine kinase whose targets are poorly understood; Akt, a serine/threonine kinase that phosphorylates the anti-apoptotic protein Bad and the ribosomal S6 kinase (S6K); GAP, GTPase activating protein; HB-EGF, heparin-binding EGF; Jak, janus kinase; PKC, protein kinase C; PLC γ , phospholipase C γ ; Shp2, Src homology domain 2-containing protein tyrosine phosphatase 2; Stat, signal transducer and activator of transcription; RAF–MEK–MAPK and PAK–JNK–JNK, two cascades of serine/threonine kinases that regulate the activity of a number of transcription factors.) Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Molecular Cell Biology 2: 127-137, copyright 2001.

Hynes, 2004). Different tyrosine phosphorylation sites in EGFR (*i.e.* Y1068 and Y1086 for Grb2 (Batzer et al., 1994), Y1148 and Y1173 for SHC (Okabayashi et al., 1994)) act as docking sites for a variety of proteins upstream of several signaling cascades, most prominently mitogen-activated protein kinase (MAPK),

phospholipase C- γ (PLC- γ), phosphoinositide 3-kinase (PI3K) and signal transducers and activators of transcription (STATs) pathways, leading to proliferation, differentiation, migration, and anti-apoptotic effects (Buettner et al., 2002; Navolanic et al., 2003; Schoeberl et al., 2002; Silva, 2004; Wells et al., 1999). Figure I.4 (Copyright Nature) shows EGFR canonical signaling pathways as defined by Yarden and Sliwkowski (Yarden & Sliwkowski, 2001).

1.1.3.1 RAS/MAPK

All EGFR family ligands and receptors are able to activate the MAPK pathway either directly through Src homology domain 2 (SH2) facilitated recruitment of growth factor receptor-bound protein 2 (Grb-2) or indirectly by phospho-tyrosine-binding (PTB) domain mediated recruitment of Grb-2 by the Src homology domain containing adaptor protein C (SHC). Recruitment of the son of sevenless protein (SOS) and other guanine nucleotide exchange factors (GEFs) occurs via the interaction of their Src homology domain 3 (SH3) with proline rich motifs in Grb-2. SOS induces Ras to exchange GDP for GTP activating it and active Ras binds and triggers – among other downstream effectors – the serine/threonine MAPK kinase kinase (MAP3K) Raf. Raf activation initiates a cascade that entails serine phosphorylation of the MAPK kinases (MAP2K) MEK1 and MEK2 and subsequent tyrosine and threonine phosphorylation of the MAPKs at the core of the pathway: the extracellular regulated kinases (ERK) 1 and 2 respectively (Chen et al., 2001; Marmor et al.,

2004; Prenzel et al., 2001). Although it is well documented that ERK threonine phosphorylation is due to MEK, the tyrosine kinase acting on ERK is yet to be identified.

Like other MAPKs, ERK1/2 are proline activated kinases and phosphorylate serine/threonine residues in the PXS/TP context (Chen et al., 2001). Among the many substrates of ERK it is possible to find many cytoplasmic, cytoskeletal, membrane bound and nuclear proteins (among them a plethora of transcription factors such Elk-1 c-Fos, c-Jun, Sp1 E2F, AP1 and the estrogen receptor (Chen et al., 1993; Chen et al., 1996; Joel et al., 1998; Marmor et al., 2004; Pende et al., 1997; Rivera et al., 1993).

Notably ERK phosphorylates several MAPK activated protein kinases (including RSKp90, RSKp70, MSK and MNK families) which control many key cellular processes and protein activity states. In fact RSK, which is best described as a multiple function ERK effector participates with ERK in the regulation of several proteins, including FOS, GSK3, MITF, ER α , ER81, tuberlin, BAD and DAPK and in a negative feedback loop to downregulate SOS. ERK and RSK are thought to control through the regulation of these proteins cell cycle progression, proliferation and cell survival (Hauge & Frodin, 2006). MSK best characterized function is to mediate ERK or p38-MAPK (p38) (a MAPK also regulated by HER related activation) regulation of gene expression by phosphorylation of transcription factors such as the cyclic AMP response binding protein (CREB) (Wiggin et al., 2002) or chromatin proteins such as Histone H3 (Soloaga et al., 2003). Finally, MNK activation by ERK or p38 results in the

phosphorylation of the eukaryotic translation initiation factor 4E (eIF4E), a key step in the initiation of protein translation (Raught & Gingras, 1999).

Membrane bound protein substrates of ERK include phospholipase A2 (PLA2) – which participates in lipid metabolism – and a negative feedback loop to downregulate EGFR (Alvarez et al., 1991; Hirabayashi et al., 2004; Lin et al., 1993). Cytoplasmic and Cytoskeletal proteins substrates of ERK include the protein adaptor paxillin, focal adhesion kinase (FAK), calpain and myosin light chain kinase (MLCK) (Huang et al., 2004). It is through its interactions with these proteins that ERK participates in the regulation of cell motility processes.

1.1.3.2 PI3K/AKT

The lipid kinase PI3K consists of two subunits, the regulatory subunit p85 and the catalytic subunit p110. Activation of the catalytic subunit results in the phosphorylation of phosphoinositide 2-phosphate (PIP2) to produce the second messenger phosphoinositide -3,4,5-phosphate (PIP3) (Bjorge et al., 1990). The effect of phosphoinositides in the cell is mediated by at least two lipid binding protein domains, namely FYVE domains which bind PIP1 and pleckstrin-homology (PH) domains which bind PIP2 and PIP3 as those in the 3'-phosphoinositide-dependent kinase 1 (PDK1) and the ACG serine/threonine kinase AKT (Blume-Jensen & Hunter, 2001).

While all EGFR family members can directly activate the MAPK pathway this is not the case for PI3K. Activation of PI3K by EGFR family members occurs

by direct phosphorylation by the receptors of the p85 regulatory subunit or by binding GTP-Ras to the p110 catalytic subunit. Since EGFR and HER2 have no binding site for the SH2- domain of the p85 subunit they need to rely on the recruitment and phosphorylation of Grb-2 associated binding protein 1 (GAB1) to trigger PI3K (Rodrigues et al., 2000), thus the activation of this kinase is relatively weak in dimers containing only these two receptors when compared to dimers containing HER3 and/or HER4 which present six and one binding motifs for the PI3K regulatory subunit respectively (Soltoff et al., 1994).

Generation of PIP2 and PIP3 on the cell membrane by PI3K results in the recruitment of AKT isoforms α , β and γ , where it is phosphorylated in its activation loop (T308 in isoform α) by PDK1 (Blume-Jensen & Hunter, 2001). Full activation of AKT kinase activity requires phosphorylation of S473 (isoform α) by the mammalian target of rapamycin (mTOR)/ricor complex (Sarbasov et al., 2005). AKT recruitment to the cell membrane is regulated by the lipid phosphatase PTEN which hydrolyzes PIP3 and PIP2 into PIP2 and PIP1 respectively (Lawlor & Alessi, 2001; Prenzel et al., 2001). Prolonged growth factor stimulation, results in nuclear translocation of AKT, where it activates several transcription factors (Meier & Hemmings, 1999). Upon its activation AKT phosphorylates several effector targets most of them within the same motif: RXRXXS/T (this motif can also be phosphorylated by MSK1) (Blume-Jensen & Hunter, 2001). It is possible to divide AKT substrates into two main groups: regulators of apoptosis/survival and regulators of cell cycle and proliferation.

AKT substrates involved in apoptosis include the pro-apoptotic protein BCL2-antagonist of cell death (BAD), several caspases (including Caspase 9) and Forkhead transcription factors. HER family mediated AKT phosphorylation of BAD at Ser 136 has been shown to be a necessary and sufficient step to block BAD induced apoptosis by dissociating the pro-apoptotic agent from BCL family members and inhibiting release of cytochrome-c from the mitochondria into the cytoplasm (Blume-Jensen et al., 1998; Datta et al., 1999; Wang et al., 1999). Cytochrome-c release into the cytoplasm results in nucleation of a protein complex known as the apoptosome which redounds in the activation of the aspartyl directed protease Caspase 9. Caspase 9 cleaves and activates Caspase 3 and Caspase 7 which in turn cleave a plethora of cellular substrates resulting in the apoptotic phenotype (Cryns & Yuan, 1998). AKT can also induce cell survival by blocking apoptosis after cytochrome release by phosphorylating Caspase 9 at S196 (Datta et al., 1999). Forkhead transcription factors activate pro-death responses by transcription of several apoptosis related genes, most prominently the activator of apoptosis Fas ligand (FasL). Phosphorylation by AKT sequesters Forkhead transcription factors in the cytoplasm, essentially shutting it down by regulating their sub-cellular localization (Biggs et al., 1999; Brunet et al., 1999).

Glycogen synthase kinase-3 (GSK-3), phosphodiesterase-3B, mTOR, insulin receptor substrate-1 (IRS-1), the Forkhead member FKHR, the cyclin-dependent kinase inhibitor p21^{CIP1/WAF1} and possibly Raf-1 are targets involved in mediating protein synthesis, glycogen metabolism and cell-cycle regulation (Blume-Jensen & Hunter, 2001). GSK-3 activity is inhibited by AKT

phosphorylation, leading to the translocation into the nucleus of β -catenin, which results in the expression of Cyclin D1 (Blume-Jensen & Hunter, 2001). Even more AKT inactivation of GSK3 results in stabilization of Cyclin D1 at the protein level by reducing phosphorylation by GSK-3 at a site that promotes Cyclin proteolysis (Diehl et al., 1998). AKT also induces survival by phosphorylating and sequestering at the cytoplasm the cyclin dependent kinases inhibitors p21, p27kip1 and p57kip2 where their activities do not have an anti-proliferative effect (Lawlor & Alessi, 2001). Finally AKT also regulates cell survival by promoting protein translation through mTOR, eIF4E and p70-S6 (Marmor et al., 2004).

1.1.3.3 PLC- γ /PKC

PLC- γ is recruited to the membrane through SH2 domain mediated binding to EGFR and HER2 or PH domain binding to PIP2 and PIP3. Subsequent RTK phosphorylation of tyrosines 771, 783 and 1254 results in PLC- γ activation. Active PLC- γ hydrolyzes PIP2 to generate the second messengers inositol 3-phosphate (IP3) and diacylglycerol (DAG) (Carpenter & Ji, 1999; Marmor et al., 2004). While phosphorylation of Y771 is dispensable for IP3 generation, phosphorylation of Y783 is essential and phosphorylation of Y1254 is necessary to achieve maximal IP3 generation (Kim et al., 1991). Binding of IP3 to its receptors (IP3R) on the endoplasmic reticulum results in the release of Ca^{2+} into the cytoplasm, thus activating a plethora of calcium/calmodulin dependent protein kinases and phosphatases. Moreover Ca^{2+} and DAG trigger the activation

of protein kinase C (PKC) catalyzing the phosphorylation of multiple substrates related to cell survival and motility (Patterson et al., 2005; Wells & Grandis, 2003). Hydrolysis of PIP2 by PLC- γ releases actin binding proteins sequestered by PIP2 such as gelsolin which cleaves current actin filaments and cofilin and profilin that serve to nucleate new actin filaments. In fact these actin binding proteins appear to regulate Arp2/3 branching at the leading edge of the cell (Wells & Grandis, 2003).

Following PIP3 hydrolysis by PLC- γ , PKC binds DAG and translocates to the membrane. Binding to DAG exposes the PKC activation loop that is phosphorylated by PDK1 at T505. Finally auto-phosphorylation of PKC c-terminal domain stabilizes it into its active form (Bornancin & Parker, 1997; Edwards & Newton, 1997). To date many substrates had been identified for PKC including substrates that interact with C-kinases (STICKs), various cytoskeletal proteins (such as tubulin or actin), scaffolding proteins (such as caveolin and A-kinase anchoring proteins (AKAPs)) and receptors for activated C-kinase (RACKs) (Jaken & Parker, 2000). Current models propose that cells express various RACKS each with a unique PKC isozyme specificity and subcellular localization thus recruiting the active isozymes to different cell compartments (Mackay & Mochly-Rosen, 2001). In the cell membrane active PKCs have been shown to phosphorylate EGFR at T654, decreasing receptor ligand affinity and receptor kinase activity (Jimenez de Asua & Goin, 1992; Morrison et al., 1993; Oliva et al., 2005)

Notably PKC- δ , a non Ca^{2+} binding isozyme, has also been found to be extensively tyrosine phosphorylated including phosphorylation at tyrosines 512 and 523 in the catalytic domain, tyrosines 52, 155 and 187 in the regulatory domain and tyrosines 311 and 332 in the hinge region (Steinberg, 2004). Of particular interest are Y332 and Y311. Y332 is flanked by a sequence region that conforms with the SH2 binding region of SHC and SHC has been shown to coordinate indirect regulatory interactions between the negative regulator of AKT, SH2 domain containing inositol phosphatase (SHIP) and PKC- δ with direct effects on cell survival (Leitges et al., 2002). Y311 is flanked by a sequence region that conforms to a SRC phosphorylation substrate. SRC phosphorylation has been linked to increased kinase activity and altered PKC- δ trafficking (Blake et al., 1999; Steinberg, 2004). PKC- δ is implicated in cell cycle arrest by promoting cyclin D and E downregulation and p27^{kip1} upregulation (Ashton et al., 1999; Fukumoto et al., 1997). It had also been implicated in transcriptional regulation by phosphorylation of STATs 1 and 3 (Jain et al., 1999; Uddin et al., 2002) and cytoskeletal reorganization, adhesion turnover and actomyosin contractility (Wells & Grandis, 2003).

1.1.3.4 STATs

Receptor phosphorylation of EGFR family members provides SH2 docking sites and promotes the recruitment of monomeric cytoplasmic signal transduction and activator of transcription proteins (STATs) (Heim et al., 1995). Receptor

bound STATs are subsequently tyrosine phosphorylated by the receptor intrinsic RTK activity or by other receptor bound kinases such as janus activated kinase (JAK) and SRC (Bowman et al., 2000). Once phosphorylated STATs dimerize via reciprocal SH2 interactions (Darnell, 1997). Within minutes of dimerization STAT dimers translocate to the nucleus where they interact with various transcriptional modulators bound to specific promoter sequences and induce gene expression (Darnell, 1997). STAT activity is ended either by de-phosphorylation or proteolytic degradation (Buettner et al., 2002; Haspel & Darnell, 1999). STATs and in particular STAT3 and STAT5 regulate the expression of many genes involved in cell cycle and survival. Several studies show that expression of the anti-apoptotic genes BCL-x_L and Mcl-1, the expression of cell cycle control genes including c-Myc and several cyclins (including cyclin D) and the expression of genes encoding the metallo-proteases MMP1, MMP2 and MMP10 and the zinc transporter LIV-1 (all genes needed for cell motility) are regulated by STAT3 and/or STAT5 (Buettner et al., 2002; Gao & Bromberg, 2006), thus STAT signaling confers EGFR family members with direct influence on cell survival, cell cycle progression and cell motility related gene transcription. Furthermore, non-tyrosine-phosphorylated, non-transcriptional STAT3 has been shown to be necessary to stabilize microtubule polymerization by binding Stathmin, a small tubulin-binding protein (Belmont & Mitchison, 1996) and tyrosine phosphorylated STAT3 was found to localize at pseudopodia of migrating cell and focal adhesions along p130Cas, FAK and Paxillin (Jia et al.,

2005; Silver et al., 2004) and to be necessary for normal p130Cas phosphorylation and cell motility in keratinocytes (Kira et al., 2002)

1.1.3.5 Other Pathways

Many other pathways are activated by or exert influence on the EGFR family. For instance HER family members might control cytoskeletal reorganization by RAS-GTP activation of the small GTPases Rho, Rac and Cdc42 (Marmor et al., 2004) and are also implicated in signaling events downstream of other cellular stimuli, including cell adhesion, lymphokines or stress signals (Carpenter, 1999). Moreover effectors of G-coupled protein receptors might require EGFR induced signaling to access the MAPK pathway (Gschwind et al., 2001; Schafer et al., 2004) and transactivation might occur via phosphorylation, autocrine ligand generation after G-coupled receptor activation, or direct phosphorylation of HER family members by non-receptor tyrosine kinases such as JAK, Pyk2, FAK and SRC, which may act downstream or upstream of the EGFR family (Fischer et al., 2003; Gschwind et al., 2001).

SRC activation upon EGF or HRG stimulation leads to the phosphorylation of additional residues in HER family members as well as cytoskeletal and focal adhesion related proteins, most notably FAK (Abram & Courtneidge, 2000; Belsches-Jablonski et al., 2001; Biscardi et al., 1999; Vadlamudi et al., 2003). Together with PI3K, PKC and ERK, SRC and FAK are the major kinases involved in cell migration, Y397 is a FAK major autophosphorylation site acting

as docking site for Src and PI3K (Cary & Guan, 1999; Eide et al., 1995; Schaller et al., 1994) and it has been shown to be necessary for both p130cas and paxillin phosphorylation in response to FAK expression (Cary & Guan, 1999; Cary et al., 1998; Schaller & Parsons, 1995; Vuori et al., 1996). In fact previous studies show that FAK Y397 is necessary for production of an invasive cellular phenotype and it is mainly through FAK localization to lamellapodia in complex with p130Cas, Src and Dock 180 that this phenotype arises (Hsia et al., 2003). Upon binding to FAK Y397 Src phosphorylates FAK Y576 and Y577 in the kinase domains increasing the enzyme activity (Calalb et al., 1995; Cary & Guan, 1999) along a number of other tyrosines which act as docking sites for paxillin p130cas and integrin β among others.

Src Y418 is a major Src autophosphorylation site, whose phosphorylation results in self-activation (although it can also be phosphorylated by RTKs) (Roskoski, 2004). Src can mediate FAK related migration by further phosphorylation of paxillin or p130Cas, which may result in focal adhesion turnover and/or lamellapodia and filipodia formation via either the RAC, JNK pathway or the RhoGAP pathway (Hsia et al., 2003; Playford & Schaller, 2004). It can also phosphorylate ERK directly or by increasing FAK activity, leading to Calpain activation and again to focal adhesion turnover and migration (Playford & Schaller, 2004).

EGFR family members can also influence migration by direct or SRC mediated phosphorylation of catenins. Catenins are known to interact with E-Cadherin, the main cell-cell adhesion protein in epithelial cells (Davis et al., 2003;

Takeichi, 1995) and regulate its function (Davis et al., 2003 JCB). E-cadherin has been identified as a tumor suppressor and its down regulation correlates with tumor cancer progression (Nollet et al., 1999; Yap, 1998). Catenin- δ and catenin- γ are members of the p120 catenin family originally identified as Src and RTK (Anastasiadis & Reynolds, 2000; Davis et al., 2003; Reynolds et al., 1992). In the absence of E-Cadherin it has been postulated that p120 catenins drive the cells towards the metastatic phenotype, by deficient regulation of several Rho-GTPase proteins (Anastasiadis et al., 2000; Anastasiadis & Reynolds, 2001). P120 Catenins had been shown to regulate E-Cadherin turnover, probably by modulating its internalization and degradation rate (Davis et al., 2003) stabilizing it on the surface when bound to it. This has direct implication on cancer metastasis as the absence or cytoplasmic localization of p120 family members lead to missregulation of Rho-GTPases and faster E-Cadherin turnover by lack of catenin dependent stabilization. HER2 overexpression in breast carcinomas inhibits the transcription of E-Cadherin (D'Souza & Taylor-Papadimitriou, 1994; Li et al., 2003), and it has also been found to destabilize the catenin-cadherin complex, leading to decreased adhesion (Jawhari et al., 1999; Li et al., 2003).

The majority of Src and RTK dependent phosphorylation sites of δ -catenin tyrosine phosphorylation are located on a relatively short region of the protein close to the N-terminus between residues 228 and 302 and EGFR has been found to directly phosphorylate p120 catenin at Y228 (Mariner et al., 2004). Moreover, SRC and EGFR phosphorylated p120 was found to localize at lamellapodia and adheren junctions and a screen of highly invasive tumor cell

lines showed high degree of Y228 phosphorylation (Mariner et al., 2004), leading to the hypothesis that Y228 phosphorylation of p120 catenin may break cell adhesion by disruption of catenin-cadherin complexes as EGFR inhibition was found to promote assembly of cell adhesions (Lorch et al., 2004).

EGFR family members are also regulated by subcellular localization. The epidermal growth factor receptor pathway 15 (EPS15) directly binds to EGFR and the clathrin adaptor protein AP-2, mediating clathrin coated pits receptor endocytosis (Torrise et al., 1999) and the E3 ubiquitin ligase Cbl targets EGFR to the lysosomes by ubiquitination of the receptor (Levkowitz et al., 1998). EGFR, HER2 and HER4 are also enriched in caveolae (Anderson, 1998; Liu et al., 2002; Schlegel et al., 2000). Caveolae are a lipid rafts rich in caveolin proteins, cholesterol and glycosphingolipids. EGFR family members interact with Caveolin-1 resulting in a decrease in kinase activity (Couet et al., 1997; Engelman et al., 1998). In fact caveolin expression is inversely correlated to HER2 signaling (Razani et al., 2001) and has been implicated in non-clathrin-mediated internalization (Pelkmans & Helenius, 2002). Furthermore, translocation of EGFR, HER3 and HER4 proteolytic fragments from early endosomes into the nucleus had been detected and it is expected that they might act directly as transcription factors adding one more way of action to EGFR family controlled sub-cellular networks (Wells & Marti, 2002).

1.1.4 Physiological Role of the EGFR Family

1.1.4.1 Heart

Embryonic lethality of HER2 (Lee et al., 1995) and HER4 (Gassmann et al., 1995) knockout mice revealed a vital role for EGFR family members in development. Mice died at post-fertilization day 10 due to abnormal cardiac and nervous system development. Failure in trabeculae – finger like extensions of the ventricular myocardium – development resulted in a mutant heart characterized by an irregular beat and reduced blood flow. HER3 knockout mice show correct trabeculae structure, but also die during gestation (day 13.5 post fertilization) due to flaws in valve formation (their atrioventricular valves are dilated and thin compared to normal mice) (Riethmacher et al., 1997). HER2 is also thought to regulate heart function in adults (Ozcelik et al., 2002), this acquires particular importance in light of the adverse cardiac related side effects that have been reported on Trastuzumab – a monoclonal anti-HER2 antibody – treatment of breast cancer patients (Slamon et al., 2001).

1.1.4.2 Nervous System

Heregulins (HRGs) – a class of ligands of that bind HER3 and HER4 – perform multiple functions in the development of the nervous system. Mutation or knockout of HER2, HER3 or heregulin (HRG/NRG1) (Garratt et al., 2000; Meyer

et al., 1997; Morris et al., 1999; Riethmacher et al., 1997; Tidcombe et al., 2003; Wolpowitz et al., 2000), resulted in the failure to form functional neuromuscular junction. Moreover, abnormal cranial and cerebellum neural architecture has been demonstrated in HER4 knockout mice (Casalini et al., 2004; Gassmann et al., 1995)), HER and HRG knockout have resulted in severe hypoplasia of the primary sympathetic ganglion chain (Britsch et al., 1998; Riethmacher et al., 1997) and complete absence of Schwann cells in peripheral nerves at late development stages (Wolpowitz et al., 2000). Finally HER2 has been shown to be necessary for schwann cell myelination (Garratt et al., 2000) and oligodendrocyte formation (Casalini et al., 2004; Kim et al., 2003) in vivo.

1.1.4.3 Epithelial development

Although EGFR plays a pivotal role in the development of epithelial cells its activity is non essential for gestation survival. Genetic knockout of EGFR in mice can result in death during gestation, birth or no later than 20 days after birth depending on mouse strain (Miettinen et al., 1995; Sibilias & Wagner, 1995; Threadgill et al., 1995). Knockout phenotypes exhibit abnormal epithelial cell proliferation, migration and differentiation in the skin, lungs and intestines, among other organs. Along with aberrant epithelial cell behavior knockout mice exhibit brain damage and those that survive birth suffer progressive neurodegeneration (Sibilias et al., 1998).

1.1.4.4 Mammary Gland

Due to the very important role EGFR family members play on mammary carcinoma, research on their physiological function in the mammary gland has been a keen focus of interest. The studies conducted to date show a very complicated picture in which all receptors of the family and the majority of their ligands play a role at different stages of development (Casalini et al., 2004). In fact, RNA amplification based analyses of the HER related ligands during mammary development show an exquisite temporal pattern of transcriptional regulation for amphiregulin (AR), betacellulin (BTC), heparin binding EGF (HB-EGF), epiregulin (EPR), EGF, HRG, and transforming growth factor α during mammary gland development, maturation and involution (Schroeder & Lee, 1998).

EGFR family members play a role in regulating growth, differentiation, apoptosis and/or architecture remodeling in the mammary gland. In the virgin gland EGFR and HER2 co-localize in all major cell types during alveolar morphogenesis, but localize different in the mature gland. Whereas EGFR and HER2 localize equally in alveoli and lactating ducts, HER3 and HER4 are preferentially localized in the alveoli (Schroeder & Lee, 1998).

RNA expression profiles assign an important role for EGFR and HER2 during mammary gland development at puberty and a minimal role during pregnancy and lactation periods. In contrast HER3 and HER4, appear to have minimal influence during puberty development of the gland and are found to be

mainly activated during pregnancy and lactation (Sebastian et al., 1998). Interestingly it has been shown that while formation of alveolar morphogenesis before pregnancy is activated via a MAPK pathway; formation of branches tubules, necessary for lactation, depends on a PI3K activated pathway (Niemann et al., 1998).

Assessment of the role of EGFR family members in immature mammary glands has been hampered by gestation lethality of the RTKs knockout. Studies in which HER4 knockout mice have been rescued from cardiac lethality (Tidcombe et al., 2003) or the roles of EGFR, HER2 and HER4 have been investigated using dominant-negative constructs of the receptors (Jones & Stern, 1999; Jones et al., 1999) show that the animals reach maturity and are fertile, but their mammary gland differentiation is aberrant and lactation is defective.

Overall, none of the EGFR family members seem to be essential per se for mammary gland development, in fact it is probable that inhibition of EGFR family member receptors together with steroid hormone receptors such as estrogen receptor is necessary for complete ablation of mammary gland development (Casalini et al., 2004; Stern, 2003).

1.1.5 EGFR Family Members Roles in Cancer

Over-expression, mutation and deregulation of EGFR family members has been correlated with cancer development and progression (Menard et al., 2004;

Mendelsohn & Baselga, 2003; Rogers et al., 2005; Ross et al., 2004a; Scagliotti et al., 2004). Moreover, deregulation of many of the signaling pathways discussed above can result in several transforming phenotypes including proliferation, migration and angiogenesis. Deregulation of the EGFR family can occur via receptor over-expression and ligand over-production by the tumor cells in an autocrine loop, paracrine growth dependent in ligand production by the surrounding stromal cells (Salomon et al., 1995) and/or constitutive receptor activation (Lonardo et al., 1990; Mendelsohn & Baselga, 2003; Pedersen et al., 2001). In this section I will present a brief summary of the EGFR family role in cancer with an emphasis on breast cancer, which is the most relevant to my work.

1.1.5.1 Breast Cancer

Over-expression of EGFR and HER2 have been related to breast cancer and correlated with poor prognosis (Citri et al., 2003; Yarden & Sliwkowski, 2001). In a recent review Ross et al., have shown that of 81 studies including 27,161 patients in 90% of the studies and in 92% of the patients HER2/neu gene amplification or HER2 protein overexpression predicted breast cancer outcome (Ross et al., 2003). However a large scale study on HER2 over-expression in breast cancer showed that only those cases in which HER2 was found to be phosphorylated predicted poor prognosis (Thor et al., 2000). In particular HER2 strongly co-dimerizes with the kinase dead HER3 (Citri et al., 2003), causing high

and sustained signaling downstream, especially through PI3K related pathways (Neve et al., 2001). In normal cells activation of HER2-HER3 is highly regulated (Citri et al., 2003) and it is considered to be the most potent signaling dimer of the EGFR family (Pinkas-Kramarski et al., 1996; Wallasch et al., 1995). HER2 overexpression increases HER3 activation in response to HRG stimulation (Aguilar et al., 1999; Lewis et al., 1996). Furthermore, various studies show that HRG activation of HER3 can induce aggressive progression of breast cancer, as indicated by anti-estrogen resistance, tumor progression, invasion and metastasis (Atlas et al., 2003; Mazumdar et al., 2001; Meiners et al., 1998; Tang et al., 1996; Tsai et al., 2000; Tsai et al., 2003; Yao et al., 2001a). The metastatic phenotype, might be due to any or a combination of the following factors: Increased cell motility, suppression of genes that keep tumor progression in check, such as p53 and PTEN, or dysregulation of oncogenic genes expression, such as SRC and AKT, and their protein activity (Stove & Bracke, 2004).

1.1.5.2 Other Cancers

EGFR overexpression and mutation play a large role in development of lung, prostate and head and neck cancer (Olapade-Olaopa et al., 2004; Rogers et al., 2005; Scagliotti et al., 2004). In particular EGFR over-expression is highly characteristic of gliomas and evidence suggests that ligand dependent activation of EGFR might play a crucial role in glioma development. In fact many tumors co-express EGFR and TGF- α or EGF (Rasheed et al., 1999). Moreover an EGFR

mutant (EGFR variant III or EGFRvIII) lacking exons 2-7 is constitutively active in a ligand independent manner (although its activity is lower than wild type (WT)) (Pedersen et al., 2001) and it has been found to localize in breast, ovary and lung carcinomas (Moscatello et al., 1995; Wong et al., 1992) as well as glioblastomas. In fact a clinical study shows that EGFRvIII overexpression in the presence of EGFR amplification is the strongest indicator for poor prognosis in glioma patients (Shinojima et al., 2003).

HER3 and/or HER4 are overexpressed in most colorectal, ovary, skin and breast tumors (Bodey et al., 1997; Marmor et al., 2004; Maurer et al., 1998; Xia et al., 1999) HER3 is related to poor prognosis in bladder and ovarian cancer as well as hepatocellular carcinomas (Chow et al., 2001; Ito et al., 2001; Simpson et al., 1995) among others and heterodimerization of this receptors with HER2 has been involved in oral squamous cell cancer and child medulloblastoma (Gilbertson et al., 1997; Xia et al., 1999). In fact, because of its slow rate of endocytosis (Baulida et al., 1996), co-expression of HER2 with EGFR, HER3 and HER4 potentiates EGFR family related transformation. HER2 has been shown to reduce the internalization and degradation rate of EGFR and increase its recycling to the cell surface after endocytosis and its ligand affinity, by stabilizing the receptor in its active state (Karunagaran et al., 1996; Worthylake et al., 1999).

The high incidence of EGFR family in cancer progression makes the understanding of its underlying cellular networks a top necessity both to obtain basic knowledge on cancer and to find new and more effective drug targets.

1.1.6 EGFR Family Related Therapeutics Compounds

A brief mention on the state of EGFR family related therapeutics is important for completion as well for its relevance to the first chapter of this thesis; by no means have I planned to be extensive in this section. More extensive reviews can be found (Fry, 2003; Heymach et al., 2006; Holbro and Hynes, 2004; Mendelshon & Baselga, 2006)

EGFR family therapeutics can be divided in two groups: one, small molecules inhibitors / ATP analogs that target the HER family catalytic domain in order to inhibit the receptors kinase activity and two, monoclonal anti-EGFR family antibodies that target the HER family members extracellular domain in order to inhibit receptor activation by blocking ligand binding and/or dimerization.

ATP competitive inhibitors have been developed for clinical applications to block access of ATP to tyrosine kinases catalytic domain (Mendelsohn & Baselga, 2003, 2006). All EGFR family effective inhibitors are derivatives of a generic quinazoline molecule (Fry, 2003) Two EGFR specific ATP analogs have been approved for the treatment of non-small-cell lung cancer and have shown to be effective against tumors that express a catalitically overactive EGFR mutant. These drugs are gefitinib (Iressa, AstraZeneca) and erlotinib (Tarceva, Genentech) (Lynch et al., 2004; Paez et al., 2004). Still in clinical trials is a group of drugs that can simultaneously block ATP binding to EGFR and HER2 such as Carnetib (CI-1033, Pfizer), EKB-569 (Wyeth-Ayerst Research) – two irreversible inhibitors - and Lapatinib (GW572016, Glaxo Smith Kline) that binds to the

inactive conformation of EGFR and HER2 kinases domain (Heymach et al., 2006; Holbro & Hynes, 2004).

The utility of monoclonal antibodies against EGFR and HER2 against tumor growth was uncovered by Greene and Mendelsohn groups respectively (Drebin et al., 1984; Sato et al., 1983). Their observation lead to the generation of several humanized anti-EGFR family antibodies for their therapeutic use. Currently approved therapeutic antibodies for treatment of colorectal cancer and neck and squamous carcinoma are cetuximab (Bristol Myers Squibb (BMS)/ImClone) which blocks ligand binding to EGFR and for breast cancer trastuzumab (Herceptin, Genentech) which inhibits HER2 activity. Still in clinical trials are pertuzumab (Omnitarg, Genentech/Roche) that targets HER2 heterodimerization, and a couple of anti-EGFR monoclonal antibodies that compete with ligand binding among them matuzumab (EMD72000, EMD/Merck KaGA) and panitumumab (ABX-EGF, Abaenix/Amaen) (Heymach et al., 2006; Holbro & Hynes, 2004).

Interestingly to the extent of my knowledge no antibodies have been raised yet that block the EGFR activation by targeting the dimerization arm in domain II or the putative dimerization sequence in domain IV. This acquires relevance, especially in autocrine systems where antibodies might not be able to compete with the local ligand concentration at the cell surface.

I.2 MASS SPECTROMETRY IN PHOSPHOPROTEOMICS

Since 1912, when Joseph John Thomson performed the first ever mass spectrometry experiment to detect the spectra of O₂, N₂, CO, CO₂ and COCl₂ (Thomson, 1913) until the late 1980s mass spectrometry application were restricted to small, thermostable products, due to the difficulty of gently ionizing and passing the ionized molecules from liquid to gas phase without excessive fragmentation (Domon & Aebersold, 2006). The inception at this time of electrospray ionization (ESI) (Fenn et al., 1989) and matrix assisted laser desorption ionization (MALDI) (Karas & Hillenkamp, 1988), two techniques capable of ionizing intact biological molecules – including polypeptides – marks an outstanding expansion in the potential of mass spectrometry applications and can be considered the starting point of the proteomics era.

1.2.1 Ion Sources

Although in this thesis work only ESI sources were used an outline of ESI and MALDI working principles are given below. For a more detailed description I recommend sections 1.6 and 1.8 of de Hoffmann and Stroobant's book (Hoffmann & Stroobant, 2001). ESI occurs when a strong electric field is applied to a liquid slowly flowing through a capillary column at atmospheric pressure. Charge accumulation is induced by the field presence at the liquid surface exiting

the capillary resulting in the formation of highly charged nano-droplets; finally the droplets are heated by passing through a curtain of hot inert gas (usually N₂) to remove the remaining solvent molecules. As the droplet charge augments or the solvent is removed the droplets charge to volume ratio increases, when columbic repulsion forces balance the cohesion forces on the droplets they break yielding two smaller droplets of equal charge and similar size, the exponential breakage of charged droplets into nano-droplets is what defines ESI (Fenn et al., 1989; Hoffmann & Stroobant, 2001).

In contrast to ESI, MALDI extracts ionized molecules from solid to gas phase. MALDI can be divided on two steps. First, in a matrix embedding step the samples are mixed in a solution of small organic molecules with high absorbance at the laser wavelength. Before analysis the mixture is completely dried, leaving a crystal where analyte molecules are completely surrounded by the matrix and isolated from each other. The second step consists in the sublimation of matrix section by quick and intense pulses of laser. Laser irradiation induces fast heating of the target zone through the energy accumulation due to the excitation of the matrix molecules. This fast heating process induces the change of the matrix phase (and embedded sample) from solid to gas, with little transfer of internal energy towards the analyte molecules which remain intact. Ionization might occur at any time during this process, but the causes of MALDI ionization are still a matter of speculation (Hoffmann & Stroobant, 2001; Karas & Hillenkamp, 1988; Zenobi & Knochenmuss, 1998)

1.2.2 Mass Spectrometers

The impact of ESI and MALDI was to catalyze the design of new mass spectrometers specifically designed to tackle questions in the newly born proteomics field (Domon & Aebersold, 2006). In this section, I will outline the function of two of these mass spectrometers describing only operation modes relevant to this thesis. For a more detailed description I recommend chapters 2 and 3 of de Hoffmann and Stroobant's book (Hoffmann & Stroobant, 2001) and review articles by Chernushevic et al. and Hopfgartner et al. (Chernushevich et al., 2001; Hopfgartner et al., 2004).

1.2.2.1 Quadrupole Time-of-Flight and IDA

Most mass spectrometry based experiments conducted during the completion of this work were done using a quadrupole time-of-flight (Qq-TOF) instrument (QSTAR-XL, Applied Biosystems) set for information (data) dependent acquisition (IDA). From now on I will use Qq-TOF or QSTAR indistinctively when referring to this type of instrument.

Figure I.5 shows the schematic of a QSTAR instrument coupled to an ESI source. The instrument consists of 3 quadrupoles, Q0, Q1 and Q2 aligned in series followed by a reflecting TOF analyzer. While both Q0 and Q2 are designed to be used in radio frequency (RF) only mode and at a pressure of several millitorr, Q1 can be used in RF only mode or radio frequency / direct current

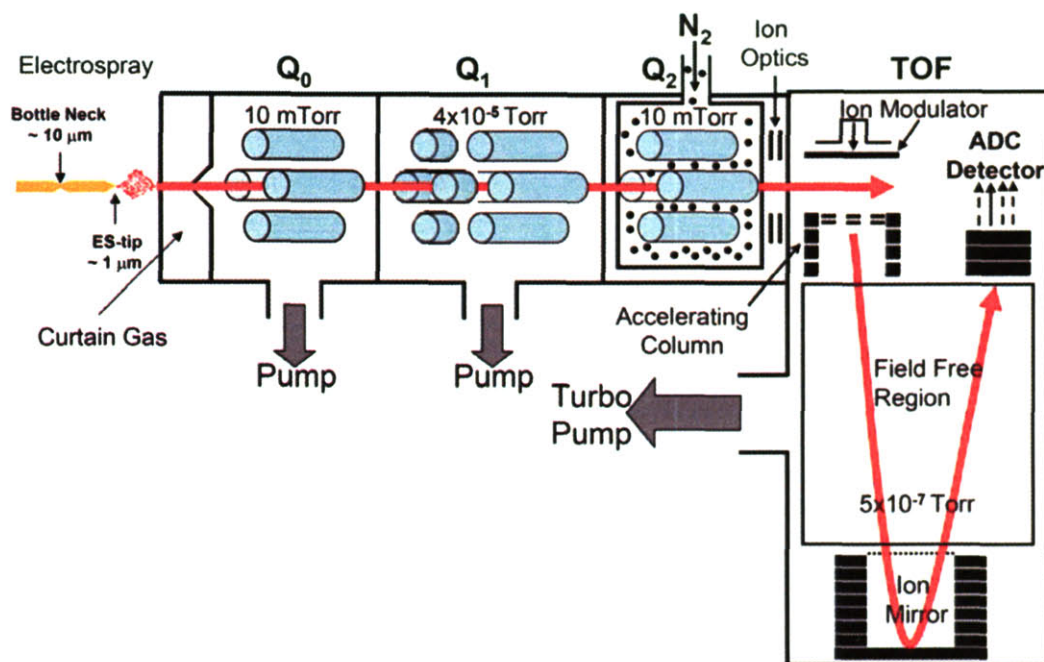


Figure I.5 – Schematic of a QSTAR instrument. The red line represents the ions path from the ESI to the detector. Adapted from (Chernushevich et al., 2001).

(RF/DC) mode and it is always kept at low pressure (order of 0.01 millitorr). The TOF analyzer is always kept close to vacuum.

In general, Q₀ is used for collisional cooling and focusing the ions coming from the ESI source into Q₁. Q₁ can be used in RF only mode or mass filter (RF/DC) mode. And Q₂ can be used as Q₀ or as a collision cell where the ions transmitted from Q₁ are broken into fragment ions by collision activated dissociation (CAD) – in which the ions are subjected to high energy collisions with an inert gas (usually Nitrogen or Argon). The resulting ions are collisionally cooled and RF focused before entering the detector.

In Q₀ and Q₂ the RF field produces a potential well that provides radial confinement of precursor and/or fragments ion according to their mass to charge

(m/z) ratio while the relative high pressure of operation ensures axial and radial collisional damping of ion movement. In the case of Q1 combination of RF/DC voltages and low pressures permits selecting a very narrow m/z window to transmit only desired ions into Q2 in what is called mass filter mode. After leaving Q2 ions are re-accelerated and focused into the TOF analyzer ion modulator. At first the ion modulator is a field free region and the ions continue their movement inside its space, but as ions get in, a pulsed high frequency (kilo-Hertz order) electric field is applied across the modulator gap and the ions are thrust perpendicular to their original direction into the analyzer accelerating column, where they reach their final kinetic energy before passing into the field free region where TOF separation occurs – at the same kinetic energy for all ions, the time it takes an ion to travel the effective length of the TOF analyzer to reach the detector is proportional to the square root of its m/z value – and ion m/z are registered by the detector and recorded using a time to digital converter (TDC) (Chernushevich et al., 2001; Hoffmann & Stroobant, 2001).

When using QSTAR mass spectrometers on IDA mode Q1 is used in mass filter mode and Q2 as a collision cell. On IDA mode the “ n ” most abundant ions (usually n is between 1 and 10) reaching the detector at a given time are identified for sequential selection in Q1 and fragmentation in Q2. The process of identifying, selecting and fragmenting, and detecting the ions fragments conforms an IDA MS cycle. The first identification step is called a full scan mass spectrum (MS event) and the identified ions are called precursor or parent ions. The selection, fragmentation and identification of ion fragments is called an MS/MS

(MS²) event. In proteomics in general an IDA MS cycle consists of one MS event followed by 5 MS/MS events (Aebersold & Mann, 2003; Chernushevich et al., 2001; Domon & Aebersold, 2006; Hoffmann & Stroobant, 2001; Zhang et al., 2005c). There are many examples of IDA applications to proteomics. Further details can be found in the following set of articles and reviews (Aebersold & Mann, 2003; Blagoev et al., 2003; Blagoev et al., 2004; Brar et al., 2006; Carr et al., 2005; Gygi & Aebersold, 2000; Gygi et al., 2000; Gygi et al., 1999; Hernandez et al., 2006; Kim et al., 2005; Kim & White, 2006; Kratchmarova et al., 2005; Loyet et al., 2005; Moser & White, 2006; Reinders & Sickmann, 2005; Schmelzle et al., 2006; Schmelzle & White, 2006; Stults & Arnott, 2005; Zhang et al., 2005c) and MSB.

For proteomics, good ion fragmentation is the key step to unequivocally identify a peptide and its post-translational modifications (PTMs). Currently most mass spectrometers use collision associated dissociation (CAD) to achieve peptide fragmentation and protein sequencing. In this technique – pioneered by the Hunt lab in the field of proteomics (Hunt et al., 1981; Hunt et al., 1986) – the ions are subjected to low energy collisions with an inert gas (usually Nitrogen or Argon) leading to the accumulation of internal energy in the peptide and it subsequent fragmentation (Hoffmann & Stroobant, 2001). Peptides subjected to CAD break across their backbone into b and y ions by severing the bond between carbonyl oxygen and amide nitrogen of two adjacent amino acids (Figure 1.6) (Roepstorff & Fohlman, 1984). If CAD fragmentation results in the detection of a heterogeneous population of b and y fragment ions, the

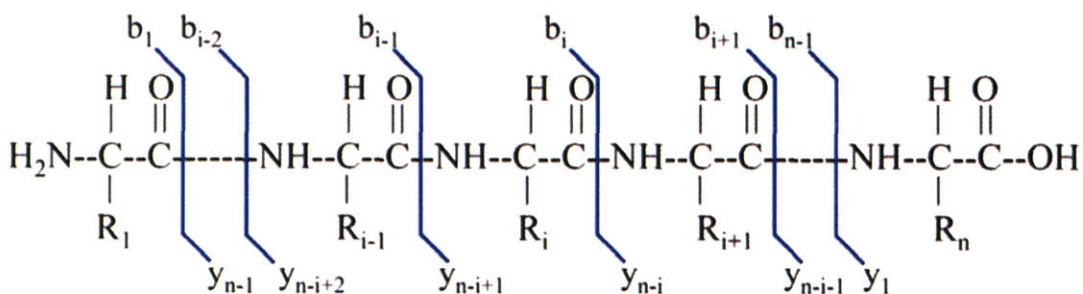


Figure I.6 – b and y ions. As a result of CAD peptides are fragmented into b and y ions by cleavage of the bonds between carbonyl oxygen and the amide nitrogen atoms.

fragmentation information and original precursor ion m/z can be used to infer the peptide sequence and PTMs. (Aebersold & Mann, 2003; Chernushevich et al., 2001; Domon & Aebersold, 2006; Hoffmann & Stroobant, 2001; Zhang et al., 2005c). One drawback of CAD is its ergodic energizing nature. CAD activates a large number of the peptide vibrational energy levels resulting primarily in the dissociation of low activation energy bonds (Cooper et al., 2005; Sleno & Volmer, 2004; Zubarev, 2004). Several PTMs, and in particular O-glycosylation, N-glycosylation and O-phosphorylation (McLafferty et al., 2001; Sleno & Volmer, 2004; Wysocki et al., 2005; Zubarev, 2004) fall in this category, thus the labile covalent bond between the modified residue and the modifying group can be broken by the internal energy transfer during CAD obscuring the identification of the PTM site and its nature. This issue has been addressed by relatively new non-ergodic fragmentation mechanisms such as electron transfer dissociation (ETD) (Syka et al., 2004) and electron capture associated dissociation (ECD) (Kelleher et al., 1999) in which the lowest activation energy states are not accessed, the PTM-residue bond is conserved and the peptide is fragmented

along its backbone into c and z ions by cleavage of the N-C_α bond (Cooper et al., 2005; Kelleher et al., 1999; Syka et al., 2004; Wysocki et al., 2005; Zubarev, 2004).

1.2.2.2 Triple Quadrupoles and MRM

The second type of mass spectrometer used during the completion of this thesis was a triple quadrupole linear ion trap (QQQ / QQ-LIT) instrument (QTRAP 4000, Applied Biosystems) set in multiple reaction monitoring mode (MRM). From now on I'll use the terms triple quad or QTRAP indistinctively when referring to this type of instrument.

Figure 1.7 shows the schematic of a QTRAP instrument. The instrument consists of 4 quadrupoles, Q0, Q1, Q2 and Q3 aligned in series where Q3 can be used as a quadrupole or as linear ion trap (LIT). Very similar to what we saw on the QSTAR; in the QTRAP, while both Q0 and Q2 are used only in radio frequency (RF) mode and at a pressure of several millitorr, Q1 and Q3 can be used in RF only or RF/DC mode and they are always kept at low pressure (order of 0.01 millitorr). Additionally Q3, but not Q1 can be used as a LIT.

In LIT mode as opposed to normal RF/DC – mass filter – mode, instead of the RF/DC potential of the quadrupole being adjusted to allow passage only of ions of selected m/z , ions with different masses are accepted into Q3 and RF voltages are adjusted to allow axial ejection of ions in an m/z selective fashion. Ejected ions are then registered and recorded by the instrument standard

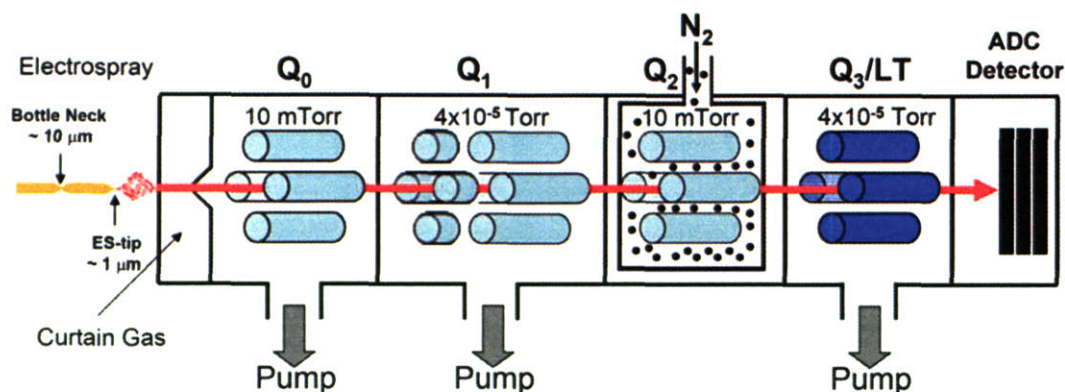


Figure 1.7 – Schematic of a QTRAP instrument. The red line represents the ions path from the ESI to the detector. Adapted from (Hopfgartner et al., 2004).

detector and analogous to digital converter (ADC). The minimal pressure in Q3 is enough to allow for collisional cooling of the trapped ions keeping their trajectories confined inside Q3 to avoid ion losses by collision with the quadrupole rods before ejection (Hoffmann & Stroobant, 2001; Hopfgartner et al., 2004).

The function and characteristics of Q0, Q1 and Q2 in the QTRAP are the same as those described for them in the QSTAR, the main difference between the instruments is the substitution of the TOF analyzer in the QSTAR by Q3 in the QTRAP. Whereas in the QSTAR all ions are passed at the same time into the TOF analyzer and m/z values are inferred from the time it takes them to reach the detector. In the QTRAP only ions that were allowed to pass through Q3 in mass filter mode or ions that were ejected in LIT mode hit the detector together. An MS/MS scan can be obtained by sequentially varying the m/z value of ions allowed through Q3 in mass filter mode or the size of ions ejected from Q3 in LIT mode (Hoffmann & Stroobant, 2001; Hopfgartner et al., 2004).

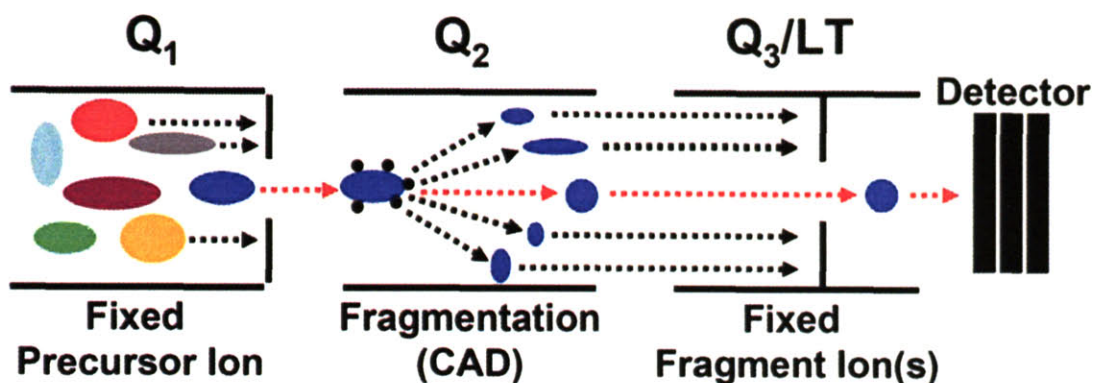


Figure I.8 – SRM/MRM Method Schematic. Multiple ions enter Q_1 , but only one/several are passed to Q_2 for CAD. All ion fragments are passed from Q_2 to Q_3 . Q_3 acts as a mass filter and only sought for ions are passed toward the detector.

When using QTRAP mass spectrometers on select or multiple reaction monitoring (SRM or MRM) mode Q_1 is used in mass filter mode, Q_2 as a collision cell and Q_3 in mass filter mode. In this way Q_1 is used to sequentially isolate desired precursor ions, Q_2 as a CAD cell to fragment these ions and Q_3 is set to let through specific fragment ions of the precursor fragmentation whose co-detection indicate the presence of a molecule of interest (Figure I.8). The only difference between SRM and MRM is that SRM isolates only one precursor ion and only one of its fragment ions gets detected, while in MRM mode it is possible to isolate many precursor ions and select several of their fragment ions for detection. In fact up to 300 fragments ions can be isolated when operating a QTRAP 4000 in MRM mode (Aebersold & Mann, 2003; Domon & Aebersold, 2006; Hoffmann & Stroobant, 2001; Hopfgartner et al., 2004).

Although not broadly used in proteomic (particularly in phosphoproteomics) MRM has been used for several small molecule applications, some examples are the quantification of DNA adducts purified from

tissue (Koc & Swenberg, 2002), the detection in biological fluids of substances whose consumption is considered doping in sports competition and thus are banned by the International Olympic Committee and/or the Horse Race Association such as beta-receptor blocking agents in human urine (Thevis et al., 2001) and anabolic steroids and other forbidden compounds in human and horse and plasma (Guan et al., 2005; Ho et al., 2006). MRM had also been used for the detection of toxic agents such as vegetal derived poisons like the pseudo-alkaloid taxine in body fluids (Frommherz et al., 2006), addictive and lethal drugs in blood for autopsies (Herrin et al., 2005) and the detection of metabolites of carcinogenic compounds related to the combustion of organic materials – usually due to smoking, grilling, or urban life – such as polycyclic aromatic hydrocarbons (PAHs) in human urine (Pigini et al., 2006).

In the proteomics field few studies have used MRM. Notably, MRM has been used in conjunction to the absolute quantification (AQUA) strategy to address the problem of absolute quantification of protein levels in biological samples and cell lysates. In AQUA for each peptide that is to be identified in MRM an equivalent synthetic, isotopically labeled peptide is added to the biological sample at a known concentration. MRM methods can detect the peptide in the biological sample and the known label. By comparing both peaks elution areas (assuming linearity between peak area and concentration of peptide) it is possible to back calculate the concentration of the peptide in the original biological sample (Kirkpatrick et al., 2005).

Other applications of MRM in proteomics has been related to the biomarkers field where it has been used find protein markers for disease severity in rheumatoid arthritis (Liao et al., 2004) and to validate a virulence biomarker for fungal pathogens (Melanson et al., 2006). MRM has also been used for low scale comparative plant proteomics (Wienkoop & Weckwerth, 2006). Finally a recently published work by Zappacosta et al.(Zappacosta et al., 2006) has applied MRM to follow the dynamic of phosphorylation of the yeast transcription factor Pho4 in response to phosphate starvation.

1.2.2.3 QSTAR v/s QTRAP: Resolution v/s Sensitivity and Dynamic Range

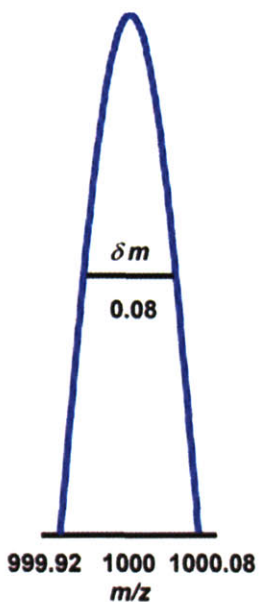


Figure I.9 – Resolution. Resolution is calculated as $R = m/\delta m$. In this case example $R = 12500$.

If I were to explain it in lay terms, it would be possible to define mass spectrometry resolution as the number characterizing how far two peaks need to be in a given mass spectra to be distinguished from each other or how sharp each individual peak is. Sharper peaks make easier to distinguish between two adjacent peaks and thus translate into better resolution.

Formally speaking, peak resolution is defined as $R = m/\delta m$, where m is the m/z value of the peak in atomic mass units (a.m.u) and δm is the full peak width at half maximum (FWHM) (Figure I.9) (Chernushevich et al., 2001; Hoffmann & Stroobant, 2001). In the case of QSTAR mass

spectrometers FWHM is usually around 10000 and goes as high as 15000 (Chernushevich et al., 2001), meaning it is possible to distinguish between peaks as close as 0.05 – 0.1 a.m.u. (1000 a.m.u). Resolution in The QTRAP is poorer with FWHM of 2500 at high resolution and FWHM of 1700 at unit resolution, meaning it is possible to distinguish between peaks as close as 0.3 – 0.5 a.m.u at high resolution and 0.5 – 0.7 at unit resolution (Hopfgartner et al., 2004).

The definitions of sensitivity and dynamic range are more intuitive and need not to be formalized. Sensitivity of an instrument is the minimal amount of sample it can detect and its dynamic range can be defined as the log of the ratio of the largest to smallest detectable signal (Hoffmann & Stroobant, 2001) and it is dependent on both the mass spectrometer's detector and ADC. Defining largest detectable signal as the number of ion counts that can be handled by the instrument before saturation of either detector or ADC and smallest detectable signal as the number of ion counts required to detect an ion above noise level (number of counts at sensitivity limit); dynamic range can be defined as $\log_{10}(\text{number of ion counts before receptor saturation} / \text{minimal number of ion counts required for detection above noise})$. In the White lab, QSTAR sensitivities can be pushed to several hundred attomoles (~100 attomole) range, while QTRAP sensitivities can be pushed down to the low attomole range (~5 attomole). The dynamic range is also better in QTRAP instrument reaching 4 – 5 logs in comparison to 2-3 log scales in the QSTAR.

The differences in resolution, sensitivity and dynamic range between QTRAP/MRM and QSTAR are mainly due to the longer scan times of the former

and better detector resolution of the second. Clearly if high speed and sensitivity is needed a QTRAP is a better instrument, but if high resolution is a must a QSTAR mass spectrometer is preferred.

1.2.3 Proteomics and Phosphoproteomics

Thanks to the steady instrumentation improvement the last decade has seen a significant increase in the efforts to use mass spectrometry to sequence the proteome and catalog the location of post-translational modification sites on proteins (Aebersold & Mann, 2003; Carr et al., 2005; Domon & Aebersold, 2006; Loyet et al., 2005; Meng et al., 2005; Reinders & Sickmann, 2005). In particular the detection – and during the last 5 years quantification – of phosphorylation sites of proteins involved in cell signaling has thrived due to the development of novel mass spectrometry and sample preparation strategies. In this section I'll address those developments with particular focus on strategies for sample separation/purification and quantification for liquid chromatography (LC) coupled mass spectrometry.

1.2.3.1 Liquid chromatography coupled mass spectrometry

Tandem liquid chromatography – electrospray ionization – mass spectrometry methods (LC-ESI-MS) take advantage of high performance liquid

chromatography (HPLC) ability to deliver a sample of interest to the ESI source during a prolonged period of time; separating the peptides that elute into the mass spectrometer over time according to their physicochemical characteristics before they enter the mass spectrometer for analysis (Aebersold & Mann, 2003; Domon & Aebersold, 2006; Hoffmann & Stroobant, 2001). LCs can be coupled to MS by using commercial available auto-samplers or by loading the sample off-line into a capillary column with an integrated electrospray tip (Davis et al., 1995; Moore et al., 1998; Zhang et al., 2005c). Many kind of beads with a diverse arrangement of physical properties can be use to pack the columns. To start the mass spectrometry analysis a liquid gradient – in which usually ionic strength or hydrophobicity increase over time – is flown through the column at low rate (less than a $\mu\text{l}/\text{minute}$). As the solvent passing through the column changes, peptides with different physicochemical characteristics elute from the column into the ESI-MS MSB (Hunt et al., 1986; Motoyama et al., 2006; Zhang et al., 2005c).

1.2.3.2 Sample Preparation for Phosphoproteomics: Before Mass Spectrometry

Due to low level and poor ionization when compared to non-phosphorylated peptides enrichment of phosphopeptides before mass spectrometry analysis is a key step in phosphoproteomics. The last years have seen the development of several novel approaches toward this end. In this section I give a brief outlook of some of these techniques.

1) Immobilized Metal Affinity Chromatography (IMAC). IMAC has been successfully used for phosphorylated peptides prior to mass spectrometry analysis (Brill et al., 2004; Ficarro et al., 2002; Salomon et al., 2003). This is an unbiased method that captures peptides phosphorylated at serine, threonine or tyrosine residues. IMAC columns are activated by using a metal cation (usually Fe^{3+}) that is retained by the column resin. When a peptide mixture is flown through an IMAC column the negatively charged phosphorylation sites are captured by the metal cation, while non-phosphorylated peptides flow through. When a complex mixture is used methyl esterification of carboxyl groups can greatly increase IMAC yields by removing non-phosphate related negative charges from the peptide mixture (Brill et al., 2004; Ficarro et al., 2002; Salomon et al., 2003).

2) Strong Cation Exchange (SCX). SCX at pH 2.7 has also been used to enrich for phosphorylated peptides, allowing the detection of more than 2000 phosphorylated peptides in the nucleus of HeLa cells. At pH 2.7 most tryptic peptides have a charge of +2 (positive charge at N terminus and lysine or arginine side chain), singly phosphorylated peptides in contrast have charge +1, doubly phosphorylated peptides are neutral, and so on. When loaded into an SCX column phosphorylated peptides will bound weaker than other peptides and will elute at low ionic strength, thus capturing early eluting fraction of SCX phosphorylation enrichment can be achieved (Beausoleil et al., 2004).

3) Biased Approaches for Specific Phosphorylation Sites.

a) Phosphoserine / Phosphothreonine. Phosphorylation in serine and threonine residues is a labile modification where the phosphate – residue bond can be easily cleaved or replaced, one strategy for enrichment takes advantage of this to replace these phosphates by biotinylated moieties that can then be used for immobilized avidin enrichment of the former phosphor-proteins from complex mixtures before mass spectrometry analysis (Oda et al., 2001). Other strategies use the same characteristic to transform phosphoserines and phosphothreonines in lysine analogs (aminoethylcysteine and β -methylaminoethylcysteine, respectively) that can be cleaved with a lysine specific protease to identify sites of phosphorylation (Knight et al., 2003; Rusnak et al., 2002).

b) Phosphotyrosine. There are several good anti-phosphotyrosine antibodies in the market that are amenable to protein immunoprecipitation (IP). This has resulted in that most strategies oriented to the identification of tyrosine phosphorylation sites have been based on protein IP with these antibodies. Phosphotyrosine IP followed in a more traditional scheme by gel electrophoresis and in gel proteolytic digestion before MS/MS analysis (Pandey et al., 2000a; Steen et al., 2001, , 2002) or in solution proteolytic digestion followed by peptide chemical modification of glutamate and aspartate for negative charge ablation on non-phosphorylated peptides and IMAC before LC–MS/MS analysis (Brill et al., 2004; Salomon et al., 2003). A novel method developed in parallel to this thesis work by Rush et al. uses anti-phosphotyrosine antibodies to IP tryptic peptides

from whole cell lysates (Rush et al., 2004) before LC–MS/MS analysis. This method advantage is that it purges the sample of most non-tyrosine-phosphorylated in the sample before mass spectrometry analysis.

1.2.3.3 Detection Strategies for Phosphoproteomics

In addition to IDA and MRM – methods used in this thesis and explained at length above – several other methods had been used to identify phosphorylation sites in mass spectrometry based proteomics in this section I give a short review of two of them.

1) Precursor ion scanning. In this mode the first analyzer (Q1 in either a Qq-TOF or triple quad) is allowed to let all peptides through for CAD in Q2. The second analyzer (TOF in Qq-TOF or Q3 in triple quad) are set to transmit only one specific fragment into the detector, which is usually a sign of a PTM-peptide. Precursor ions that generate this fragment are then isolated in Q1, fragmented in Q2 and analyzed in the second analyzer to obtain the peptide complete sequence and site of PTM (Domon & Aebersold, 2006). This setting has been used in both negative and positive ion mode to identify tyrosine phosphorylated peptides (Hinsby et al., 2003; Steen et al., 2003).

2) Neutral loss scanning. In this mode the first and second analyzer are synchronized so that the m/z difference of ions passing through them is kept

constant all the time. The mass difference corresponds to a neutral loss from the ion scanned in the first analyzer that was lost during CAD in the collision cell. When such event occurs the precursor ion is isolated in Q1, fragmented in Q2 and analyzed in the second analyzer to obtain the complete peptide sequence and site of PTM if applicable (Domon & Aebersold, 2006). This mode had been use in phosphoproteomics to identify sites of threonine or serine phosphorylation which present a neutral loss of 98, 49, 33.7, 24.5, etc ($z= 1,2,3,4$, etc. respectively) (Bateman et al., 2002; Schlosser et al., 2001)

1.2.3.4 Quantification Methods for Mass Spectrometry

Labeling strategies for mass spectrometry based quantification can be divided essentially in three groups: metabolic stable isotope labeling, isotope tagging by chemical reaction and stable isotope incorporation via enzyme reaction (Aebersold & Mann, 2003). The first group main strategy is stable isotope labeling with amino acids in cell culture (SILAC) (Ong et al., 2002), but metabolic labeling has also been achieved using heavy salts in cell culture (Conrads et al., 2002). The second group is more diverse including chemistries targeted to label sulphydryl groups (isotope coded affinity tagging – ICAT) (Gygi et al., 1999; Zhou et al., 2002), phosphate ester groups (Oda et al., 2001; Zhou et al., 2001), N-linked carbohydrates (Zhang et al., 2003; Zhang et al., 2005a), serine and cysteine hydrolases active groups (Greenbaum et al., 2000; Liu et al., 1999) and chemistries targeted to label amine groups (Munchbach et al., 2000)

the main exponent of this group are iTRAQ reagents whose application was first reported by Ross et al. (Ross et al., 2004b). Finally the third strategy, enzymatic labeling is achieved by transferring ^{18}O from water to peptides during enzymatic proteolysis (Mirgorodskaya et al., 2005; Mirgorodskaya et al., 2000; Reynolds et al., 2002; Yao et al., 2001b).

In this section I will briefly review and discuss the advantages and disadvantages of SILAC and iTRAQ labeling techniques for mass spectrometry quantification. These techniques are the most amenable and widely use techniques for phosphoproteomics quantification. As a caution note it has to be said that both techniques are used to measure changes and not absolute levels of phosphorylation. For absolute phosphorylation measurements the AQUA method (Kirkpatrick et al., 2005) outlined in the triple quadrupole MRM section is recommended.

1) SILAC is a technique that relies in the use of cell culture media containing exclusive combinations of ^{12}C , ^{13}C , ^{14}N and ^{15}N arginine and/or lysine for quantification in MS to distinguish proteins from cells grown using the different media. In theory 100% of the proteins in a cell culture could be labeled using this technique and usually after 5 media changes from normal to SILAC media protein labeling in a culture is virtually complete. SILAC quantification occurs in MS mode (Ong et al., 2002).

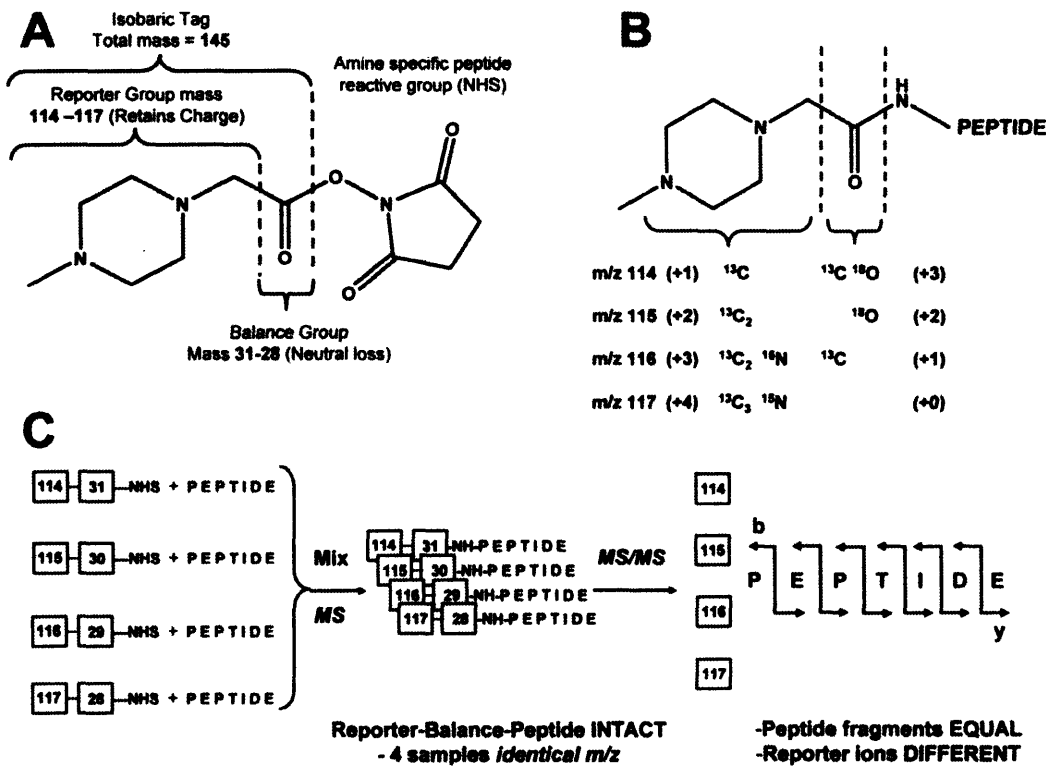


Figure I.10 – iTRAQ Chemistry. (A), diagram showing the components of the multiplexed isobaric tagging chemistry. The complete molecule consists of a reporter group (based on *N*-methylpiperazine), a mass balance group (carbonyl), and a peptide-reactive group (NHS ester). The overall mass of reporter and balance components of the molecule are kept constant using differential isotopic enrichment with ^{13}C , ^{15}N , and ^{18}O atoms (B), thus avoiding problems with chromatographic separation seen with enrichment involving deuterium substitution. The number and position of enriched centers in the ring has no effect on chromatographic or MS behavior. The reporter group ranges in mass from m/z 114.1 to 117.1, while the balance group ranges in mass from 28 to 31 Da, such that the combined mass remains constant (145.1 Da) for each of the four reagents. (B) when reacted with a peptide, the tag forms an amide linkage to any peptide amine (N-terminal or α amino group of lysine). These amide linkages fragment in a similar fashion to backbone peptide bonds when subjected to CID. Following fragmentation of the tag amide bond, however, the balance (carbonyl) moiety is lost (neutral loss), while charge is retained by the reporter group fragment. The numbers in parentheses indicate the number of enriched centers in each section of the molecule. (C) illustration of the isotopic tagging used to arrive at four isobaric combinations with four different reporter group masses. A mixture of four identical peptides each labeled with one member of the multiplex set appears as a single, unresolved precursor ion in MS (identical m/z). Following CID, the four reporter group ions appear as distinct masses (114–117 Da). All other sequence-informative fragment ions (b-, y-, etc.) remain isobaric, and their individual ion current signals (signal intensities) are additive. This remains the case even for those tryptic peptides that are labeled at both the N terminus and lysine side chains, and those peptides containing internal lysine residues due to incomplete cleavage with trypsin. The relative concentration of the peptides is thus deduced from the relative intensities of the corresponding reporter ions. In contrast to ICAT and similar mass-difference labeling strategies, quantitation is thus performed at the MS/MS stage rather than in MS. Reprinted by permission of Figure I.10 was Reprinted by permission from ASBMB Journals: *Molecular Cellular Proteomics* 3: 1154-1169, copyright 2004

2) iTRAQ labeling of whole proteins or tryptic peptides is a chemical modification to peptides/proteins that uses four amine reactive isobaric tags capable of covalently binding to a peptide N-terminus and to lysine side chains and differ only in the positioning of isotopically tagged atoms (Figure 1.10, Copyright Molecular Cellular Proteomics). iTRAQ labeling is always done after cell lysis and protein digestion and can be used to quantify up to four samples at a time. iTRAQ quantification occurs in MS/MS mode by the detection of distinctive marker ions of m/z 114 – 117 that are released when labeled peptides are subjected to CAD (Ross et al., 2004b).

While the advantage of SILAC is labeling during cell culture, avoiding further chemical modification of the sample and allowing the use of as much sample as needed. iTRAQ labeling gives the possibility of studying both experimental and clinical samples, such as tumor biopsies. Also, the MS/MS quantification gives a cleaner mass spectrometry run allowing for easier study of complex mixtures and it is easily applicable for site specific quantification of phosphorylation in contrast to SILAC which has been usually used to measure total protein phosphorylation. With their merits and de-merits SILAC and iTRAQ had been successfully used to study epidermal growth factor receptor (EGFR) mediated signaling (Blagoev et al., 2004) and to compare EGFR and platelet derived growth factor receptor (PDGFR) mediated signaling (Kratchmarova et al., 2005) (SILAC) and EGFR family mediated signaling (Zhang et al., 2005c) (MSB), CD3 and CD28 activated pathways (Kim & White, 2006) and insulin related signaling (Schmelzle et al., 2006) (iTRAQ).

I.3 MOTIVATION

All phosphorylation-mediated cellular signaling cascades are bound by the same principles. After receptor-ligand stimulation a dynamic relationship between component proteins in the pathway generates site and time-specific phosphorylation/dephosphorylation events that propagate down the cascade until the desired response is elicited. Most proteins have more than one potential phosphorylation site, and phosphorylation/dephosphorylation of different sites in the same protein may lead to different responses; *i.e.* activation events may involve a particular protein, but specific phosphorylation sites regulate the cellular activity (Gotoh et al., 1997; Moro et al., 2002). So far, most of the literature deals with either time dynamics on the activation of a handful of proteins and phosphorylation sites (Irish et al., 2004; Krutzik et al., 2004) or global identification of protein phosphorylation sites under static conditions (Beausoleil et al., 2004; Brill et al., 2004; Ficarro et al., 2002; Salomon et al., 2003).

Our current knowledge about the EGFR-related pathways dynamics and the phosphorylation sites of their proteins represent the summary of decades of work by many groups using traditional biochemistry tools to study the activation of only several proteins at a time. Only recently has mass spectrometry been used to identify many tyrosine phosphorylation sites in receptor tyrosine kinase pathways (Hinsby et al., 2003; Pandey et al., 2002; Pandey et al., 2000b; Steen et al., 2002) and to monitor total phosphotyrosine content in pathway proteins

over time (Blagoev et al., 2004). However, these studies have not provided information on the dynamic regulation of specific tyrosine phosphorylation sites.

This thesis is divided in two sections. In the first section I will talk of a proof of principle therapeutic approach developed in collaboration between professors Lauffenburger and Wittrup laboratories to block EGFR activation without necessarily blocking ligand binding. This approach has the potential to be highly useful in EGFR driven cancers, especially in those where intracellular signaling potentiated by autocrine loops and ligand binding blocking antibodies seem not to work.

The second section will deal with the development, application to EGFR/HER2 signal and further refinement of a new highly quantitative and specific mass spectrometry based technology to study cellular tyrosine phosphorylation networks with temporal and biological condition resolution that was developed in collaboration between professors Lauffenburger and White laboratories. This technology allows the simultaneous discovery, monitoring and measurement across time and biological condition space of tyrosine phosphorylation events with high accuracy, expanding our knowledge of cellular networks and generating high quality data that can be used for pathway modeling, detection of non-intuitive pathway biomarkers, the generation of new hypothesis on the role of different proteins within the cellular environment and their role in cell phenotypic responses and possibly the discovery of new and more effective drugs targets for cancer.

I.4 REFERENCES

Abram CL, Courtneidge SA (2000) Src family tyrosine kinases and growth factor signaling. *Exp Cell Res* 254: 1-13.

Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. *Nature* 422: 198-207.

Aguilar Z, Akita RW, Finn RS, Ramos BL, Pegram MD, Kabbinavar FF, Pietras RJ, Pisacane P, Sliwkowski MX, Slamon DJ (1999) Biologic effects of heregulin/neu differentiation factor on normal and malignant human breast and ovarian epithelial cells. *Oncogene* 18: 6050-6062.

Alvarez E, Northwood IC, Gonzalez FA, Latour DA, Seth A, Abate C, Curran T, Davis RJ (1991) Pro-Leu-Ser/Thr-Pro is a consensus primary sequence for substrate protein phosphorylation. Characterization of the phosphorylation of c-myc and c-jun proteins by an epidermal growth factor receptor threonine 669 protein kinase. *J Biol Chem* 266: 15277-15285.

Anastasiadis PZ, Moon SY, Thoreson MA, Mariner DJ, Crawford HC, Zheng Y, Reynolds AB (2000) Inhibition of RhoA by p120 catenin. *Nat Cell Biol* 2: 637-644.

Anastasiadis PZ, Reynolds AB (2000) The p120 catenin family: complex roles in adhesion, signaling and cancer. *J Cell Sci* 113 (Pt 8): 1319-1334.

Anastasiadis PZ, Reynolds AB (2001) Regulation of Rho GTPases by p120-catenin. *Curr Opin Cell Biol* 13: 604-610.

Anderson RG (1998) The caveolae membrane system. *Annu Rev Biochem* 67: 199-225.

Ashton AW, Watanabe G, Albanese C, Harrington EO, Ware JA, Pestell RG (1999) Protein kinase Cdelta inhibition of S-phase transition in capillary endothelial cells involves the cyclin-dependent kinase inhibitor p27(Kip1). *J Biol Chem* 274: 20805-20811.

Atlas E, Cardillo M, Mehmi I, Zahedkargaran H, Tang C, Lupu R (2003) Heregulin is sufficient for the promotion of tumorigenicity and metastasis of breast cancer cells in vivo. *Mol Cancer Res* 1: 165-175.

Bateman RH, Carruthers R, Hoyes JB, Jones C, Langridge JI, Millar A, Vissers JP (2002) A novel precursor ion discovery method on a hybrid quadrupole orthogonal acceleration time-of-flight (Q-TOF) mass spectrometer for studying protein phosphorylation. *J Am Soc Mass Spectrom* 13: 792-803.

Batzer AG, Rotin D, Urena JM, Skolnik EY, Schlessinger J (1994) Hierarchy of binding sites for Grb2 and Shc on the epidermal growth factor receptor. *Mol Cell Biol* 14: 5192-5201.

Baulida J, Kraus MH, Alimandi M, Di Fiore PP, Carpenter G (1996) All ErbB receptors other than the epidermal growth factor receptor are endocytosis impaired. *J Biol Chem* 271: 5251-5257.

Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villen J, Li J, Cohn MA, Cantley LC, Gygi SP (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A* 101: 12130-12135.

Beerli RR, Hynes NE (1996) Epidermal growth factor-related peptides activate distinct subsets of ErbB receptors and differ in their biological activities. *J Biol Chem* 271: 6071-6076.

Belmont LD, Mitchison TJ (1996) Identification of a protein that interacts with tubulin dimers and increases the catastrophe rate of microtubules. *Cell* 84: 623-631.

Belsches-Jablonski AP, Biscardi JS, Peavy DR, Tice DA, Romney DA, Parsons SJ (2001) Src family kinases and HER2 interactions in human breast cancer cell growth and survival. *Oncogene* 20: 1465-1475.

Berezov A, Chen J, Liu Q, Zhang HT, Greene MI, Murali R (2002) Disabling receptor ensembles with rationally designed interface peptidomimetics. *J Biol Chem*: Epub 2002 May 2014.

Biggs WH, 3rd, Meisenhelder J, Hunter T, Cavenee WK, Arden KC (1999) Protein kinase B/Akt-mediated phosphorylation promotes nuclear exclusion of

the winged helix transcription factor FKHR1. *Proc Natl Acad Sci U S A* 96: 7421-7426.

Biscardi JS, Tice DA, Parsons SJ (1999) c-Src, receptor tyrosine kinases, and human cancer. *Adv Cancer Res* 76: 61-119.

Bjorge JD, Chan TO, Antczak M, Kung HJ, Fujita DJ (1990) Activated type I phosphatidylinositol kinase is associated with the epidermal growth factor (EGF) receptor following EGF stimulation. *Proc Natl Acad Sci U S A* 87: 3816-3820.

Blagoev B, Kratchmarova I, Ong SE, Nielsen M, Foster LJ, Mann M (2003) A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nat Biotechnol* 21: 315-318.

Blagoev B, Ong SE, Kratchmarova I, Mann M (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat Biotechnol* 22: 1139-1145.

Blake RA, Garcia-Paramio P, Parker PJ, Courtneidge SA (1999) Src promotes PKCdelta degradation. *Cell Growth Differ* 10: 231-241.

Blume-Jensen P, Hunter T (2001) Oncogenic kinase signalling. *Nature* 411: 355-365.

Blume-Jensen P, Janknecht R, Hunter T (1998) The kit receptor promotes cell survival via activation of PI 3-kinase and subsequent Akt-mediated phosphorylation of Bad on Ser136. *Curr Biol* 8: 779-782.

Bodey B, Bodey B, Jr., Groger AM, Luck JV, Siegel SE, Taylor CR, Kaiser HE (1997) Clinical and prognostic significance of the expression of the c-erbB-2 and c-erbB-3 oncoproteins in primary and metastatic malignant melanomas and breast carcinomas. *Anticancer Res* 17: 1319-1330.

Boni-Schnetzler M, Pilch PF (1987) Mechanism of epidermal growth factor receptor autophosphorylation and high-affinity binding. *Proc Natl Acad Sci U S A* 84: 7832-7836.

Bornancin F, Parker PJ (1997) Phosphorylation of protein kinase C-alpha on serine 657 controls the accumulation of active enzyme and contributes to its phosphatase-resistant state. *J Biol Chem* 272: 3544-3549.

Bouyain S, Longo PA, Li S, Ferguson KM, Leahy DJ (2005) The extracellular region of ErbB4 adopts a tethered conformation in the absence of ligand. *Proc Natl Acad Sci U S A*: Epub 2005 Oct 2003.

Bowman T, Garcia R, Turkson J, Jove R (2000) STATs in oncogenesis. *Oncogene* 19: 2474-2488.

Brar GA, Kiburz BM, Zhang Y, Kim JE, White F, Amon A (2006) Rec8 phosphorylation and recombination promote the step-wise loss of cohesins in meiosis. *Nature*: Epub 2006 May 2003.

Brill LM, Salomon AR, Ficarro SB, Mukherji M, Stettler-Gill M, Peters EC (2004) Robust phosphoproteomic profiling of tyrosine phosphorylation sites from human T cells using immobilized metal affinity chromatography and tandem mass spectrometry. *Anal Chem* 76: 2763-2772.

Britsch S, Li L, Kirchhoff S, Theuring F, Brinkmann V, Birchmeier C, Riethmacher D (1998) The ErbB2 and ErbB3 receptors and their ligand, neuregulin-1, are essential for development of the sympathetic nervous system. *Genes Dev* 12: 1825-1836.

Brunet A, Bonni A, Zigmund MJ, Lin MZ, Joo P, Hu LS, Anderson MJ, Arden KC, Blenis J, Greenberg ME (1999) Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell* 96: 857-868.

Buettner R, Mora LB, Jove R (2002) Activated STAT signaling in human tumors provides novel molecular targets for therapeutic intervention. *Clin Cancer Res* 8: 945-954.

Burgess AW, Cho HS, Eigenbrot C, Ferguson KM, Garrett TP, Leahy DJ, Lemmon MA, Sliwkowski MX, Ward CW, Yokoyama S (2003) An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Mol Cell* 12: 541-552.

Calalb MB, Polte TR, Hanks SK (1995) Tyrosine phosphorylation of focal adhesion kinase at sites in the catalytic domain regulates kinase activity: a role for Src family kinases. *Mol Cell Biol* 15: 954-963.

Carpenter G (1999) Employment of the epidermal growth factor receptor in growth factor-independent signaling pathways. *J Cell Biol* 146: 697-702.

Carpenter G, Ji Q (1999) Phospholipase C-gamma as a signal-transducing element. *Exp Cell Res* 253: 15-24.

Carr SA, Annan RS, Huddleston MJ (2005) Mapping posttranslational modifications of proteins by MS-based selective detection: application to phosphoproteomics. *Methods Enzymol* 405: 82-115.

Cary LA, Guan JL (1999) Focal adhesion kinase in integrin-mediated signaling. *Front Biosci* 4: D102-113.

Cary LA, Han DC, Polte TR, Hanks SK, Guan JL (1998) Identification of p130Cas as a mediator of focal adhesion kinase-promoted cell migration. *J Cell Biol* 140: 211-221.

Casalini P, Iorio MV, Galmozzi E, Menard S (2004) Role of HER receptors family in development and differentiation. *J Cell Physiol* 200: 343-350.

Chen RH, Abate C, Blenis J (1993) Phosphorylation of the c-Fos transrepression domain by mitogen-activated protein kinase and 90-kDa ribosomal S6 kinase. *Proc Natl Acad Sci U S A* 90: 10952-10956.

Chen RH, Juo PC, Curran T, Blenis J (1996) Phosphorylation of c-Fos at the C-terminus enhances its transforming activity. *Oncogene* 12: 1493-1502.

Chen Z, Gibson TB, Robinson F, Silvestro L, Pearson G, Xu BE, Wright A, Vanderbilt C, Cobb MH (2001) MAP kinases. *Chemical Reviews* 101: 2449-2476.

Chernushevich IV, Loboda AV, Thomson BA (2001) An introduction to quadrupole-time-of-flight mass spectrometry. *J Mass Spectrom* 36: 849-865.

Cho HS, Leahy DJ (2002) Structure of the extracellular region of HER3 reveals an interdomain tether. *Science* 297: 1330-1333.

Cho HS, Mason K, Ramyar KX, Stanley AM, Gabelli SB, Denney DW, Jr., Leahy DJ (2003) Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. *Nature* 421: 756-760.

Chow NH, Chan SH, Tzai TS, Ho CL, Liu HS (2001) Expression profiles of ErbB family receptors and prognosis in primary transitional cell carcinoma of the urinary bladder. *Clin Cancer Res* 7: 1957-1962.

Citri A, Skaria KB, Yarden Y (2003) The deaf and the dumb: the biology of ErbB-2 and ErbB-3. *Exp Cell Res* 284: 54-65.

Citri A, Yarden Y (2006) EGF-ERBB signalling: towards the systems level. *Nat Rev Mol Cell Biol* 7: 505-516.

Cohen S (1965) The stimulation of epidermal proliferation by a specific protein (EGF). *Dev Biol* 12: 394-407.

Cohen S, Carpenter G (1975) Human epidermal growth factor: isolation and chemical and biological properties. *Proc Natl Acad Sci U S A* 72: 1317-1321.

Cohen S, Elliott GA (1963) The stimulation of epidermal keratinization by a protein isolated from the submaxillary gland of the mouse. *J Invest Dermatol* 40: 1-5.

Cohen S, Fava RA, Sawyer ST (1982) Purification and characterization of epidermal growth factor receptor/protein kinase from normal mouse liver. *Proc Natl Acad Sci U S A* 79: 6237-6241.

Conrads TP, Issaq HJ, Veenstra TD (2002) New tools for quantitative phosphoproteome analysis. *Biochem Biophys Res Commun* 290: 885-890.

Cooper HJ, Hakansson K, Marshall AG (2005) The role of electron capture dissociation in biomolecular analysis. *Mass Spectrom Rev* 24: 201-222.

Couet J, Sargiacomo M, Lisanti MP (1997) Interaction of a receptor tyrosine kinase, EGF-R, with caveolins. Caveolin binding negatively regulates tyrosine and serine/threonine kinase activities. *J Biol Chem* 272: 30429-30438.

Cryns V, Yuan J (1998) Proteases to die for. *Genes Dev* 12: 1551-1570.

D'Souza B, Taylor-Papadimitriou J (1994) Overexpression of ERBB2 in human mammary epithelial cells signals inhibition of transcription of the E-cadherin gene. *Proc Natl Acad Sci U S A* 91: 7202-7206.

Darnell JE, Jr. (1997) STATs and gene regulation. *Science* 277: 1630-1635.

Datta SR, Brunet A, Greenberg ME (1999) Cellular survival: a play in three Akts. *Genes Dev* 13: 2905-2927.

Davis MA, Ireton RC, Reynolds AB (2003) A core function for p120-catenin in cadherin turnover. *J Cell Biol* 163: 525-534.

Davis MT, Stahl DC, Hefta SA, Lee TD (1995) A microscale electrospray interface for on-line, capillary liquid chromatography/tandem mass spectrometry of complex peptide mixtures. *Anal Chem* 67: 4549-4556.

Dawson JP, Berger MB, Lin CC, Schlessinger J, Lemmon MA, Ferguson KM (2005) Epidermal growth factor receptor dimerization and activation require ligand-induced conformational changes in the dimer interface. *Mol Cell Biol* 25: 7734-7742.

Diehl JA, Cheng M, Roussel MF, Sherr CJ (1998) Glycogen synthase kinase-3beta regulates cyclin D1 proteolysis and subcellular localization. *Genes Dev* 12: 3499-3511.

Domon B, Aebersold R (2006) Mass spectrometry and protein analysis. *Science* 312: 212-217.

Drebin JA, Stern DF, Link VC, Weinberg RA, Greene MI (1984) Monoclonal antibodies identify a cell-surface antigen associated with an activated cellular oncogene. *Nature* 312: 545-548.

Edwards AS, Newton AC (1997) Phosphorylation at conserved carboxyl-terminal hydrophobic motif regulates the catalytic and regulatory domains of protein kinase C. *J Biol Chem* 272: 18382-18390.

Eide BL, Turck CW, Escobedo JA (1995) Identification of Tyr-397 as the primary site of tyrosine phosphorylation and pp60src association in the focal adhesion kinase, pp125FAK. *Mol Cell Biol* 15: 2819-2827.

Elenius K, Paul S, Allison G, Sun J, Klagsbrun M (1997) Activation of HER4 by heparin-binding EGF-like growth factor stimulates chemotaxis but not proliferation. *Embo J* 16: 1268-1278.

Engelman JA, Lee RJ, Karnezis A, Bearss DJ, Webster M, Siegel P, Muller WJ, Windle JJ, Pestell RG, Lisanti MP (1998) Reciprocal regulation of neu tyrosine kinase activity and caveolin-1 protein expression in vitro and in vivo. Implications for human breast cancer. *J Biol Chem* 273: 20448-20455.

Falls DL (2003) Neuregulins: functions, forms, and signaling strategies. *Exp Cell Res* 284: 14-30.

Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246: 64-71.

Ferguson KM, Berger MB, Mendrola JM, Cho HS, Leahy DJ, Lemmon MA (2003) EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization. *Mol Cell* 11: 507-517.

Ficarro SB, McClelland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM (2002) Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotechnol* 20: 301-305.

Fischer OM, Hart S, Gschwind A, Ullrich A (2003) EGFR signal transactivation in cancer cells. *Biochem Soc Trans* 31: 1203-1208.

Frommherz L, Kintz P, Kijewski H, Kohler H, Lehr M, Brinkmann B, Beike J (2006) Quantitative determination of taxine B in body fluids by LC-MS-MS. *Int J Legal Med* 6: 1-6.

Fry DW (2003) Mechanism of action of erbB tyrosine kinase inhibitors. *Exp Cell Res* 284: 131-139.

Fukumoto S, Nishizawa Y, Hosoi M, Koyama H, Yamakawa K, Ohno S, Morii H (1997) Protein kinase C delta inhibits the proliferation of vascular smooth muscle cells by suppressing G1 cyclin expression. *J Biol Chem* 272: 13816-13822.

Gadella TW, Jr., Jovin TM (1995) Oligomerization of epidermal growth factor receptors on A431 cells studied by time-resolved fluorescence imaging microscopy. A stereochemical model for tyrosine kinase receptor activation. *J Cell Biol* 129: 1543-1558.

Gao SP, Bromberg JF (2006) Touched and moved by STAT3. *Sci STKE* 11.

Garratt AN, Voiculescu O, Topilko P, Charnay P, Birchmeier C (2000) A dual role of erbB2 in myelination and in expansion of the schwann cell precursor pool. *J Cell Biol* 148: 1035-1046.

Garrett TP, McKern NM, Lou M, Elleman TC, Adams TE, Lovrecz GO, Kofler M, Jorissen RN, Nice EC, Burgess AW, Ward CW (2003) The crystal structure of a truncated ErbB2 ectodomain reveals an active conformation, poised to interact with other ErbB receptors. *Mol Cell* 11: 495-505.

Garrett TP, McKern NM, Lou M, Elleman TC, Adams TE, Lovrecz GO, Zhu HJ, Walker F, Frenkel MJ, Hoyne PA, Jorissen RN, Nice EC, Burgess AW, Ward CW (2002) Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor alpha. *Cell* 110: 763-773.

Gassmann M, Casagrande F, Orioli D, Simon H, Lai C, Klein R, Lemke G (1995) Aberrant neural and cardiac development in mice lacking the ErbB4 neuregulin receptor. *Nature* 378: 390-394.

Gilbertson RJ, Perry RH, Kelly PJ, Pearson AD, Lunec J (1997) Prognostic significance of HER2 and HER4 coexpression in childhood medulloblastoma. *Cancer Res* 57: 3272-3280.

Gotoh N, Toyoda M, Shibuya M (1997) Tyrosine phosphorylation sites at amino acids 239 and 240 of Shc are involved in epidermal growth factor-induced

mitogenic signaling that is distinct from Ras/mitogen-activated protein kinase activation. *Mol Cell Biol* 17: 1824-1831.

Greenbaum D, Medzihradzky KF, Burlingame A, Bogyo M (2000) Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chem Biol* 7: 569-581.

Gschwind A, Zwick E, Prenzel N, Leserer M, Ullrich A (2001) Cell communication networks: epidermal growth factor receptor transactivation as the paradigm for interreceptor signal transmission. *Oncogene* 20: 1594-1600.

Guan F, Uboh CE, Soma LR, Luo Y, Rudy J, Tobin T (2005) Detection, quantification and confirmation of anabolic steroids in equine plasma by liquid chromatography and tandem mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* 829: 56-68.

Guy PM, Platko JV, Cantley LC, Cerione RA, Carraway KL, 3rd (1994) Insect cell-expressed p180erbB3 possesses an impaired tyrosine kinase activity. *Proc Natl Acad Sci U S A* 91: 8132-8136.

Gygi SP, Aebersold R (2000) Mass spectrometry and proteomics. *Curr Opin Chem Biol* 4: 489-494.

Gygi SP, Rist B, Aebersold R (2000) Measuring gene expression by quantitative proteome analysis. *Curr Opin Biotechnol* 11: 396-401.

Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 17: 994-999.

Harris RC, Chung E, Coffey RJ (2003) EGF receptor ligands. *Exp Cell Res* 284: 2-13.

Haspel RL, Darnell JE, Jr. (1999) A nuclear protein tyrosine phosphatase is required for the inactivation of Stat1. *Proc Natl Acad Sci U S A* 96: 10188-10193.

Hauge C, Frodin M (2006) RSK and MSK in MAP kinase signalling. *J Cell Sci* 119: 3021-3023.

Heim MH, Kerr IM, Stark GR, Darnell JE, Jr. (1995) Contribution of STAT SH2 groups to specific interferon signaling by the Jak-STAT pathway. *Science* 267: 1347-1349.

Hernandez P, Muller M, Appel RD (2006) Automated protein identification by tandem mass spectrometry: issues and strategies. *Mass Spectrom Rev* 25: 235-254.

Herrin GL, McCurdy HH, Wall WH (2005) Investigation of an LC-MS-MS (QTrap) method for the rapid screening and identification of drugs in postmortem toxicology whole blood samples. *J Anal Toxicol* 29: 599-606.

Heymach JV, Nilsson M, Blumenschein G, Papadimitrakopoulou V, Herbst R (2006) Epidermal growth factor receptor inhibitors in development for the treatment of non-small cell lung cancer. *Clin Cancer Res* 12: 4441s-4445s.

Hinsby AM, Olsen JV, Bennett KL, Mann M (2003) Signaling initiated by overexpression of the fibroblast growth factor receptor-1 investigated by mass spectrometry. *Mol Cell Proteomics* 2: 29-36.

Hirabayashi T, Murayama T, Shimizu T (2004) Regulatory mechanism and physiological role of cytosolic phospholipase A2. *Biol Pharm Bull* 27: 1168-1173.

Ho EN, Leung DK, Wan TS, Yu NH (2006) Comprehensive screening of anabolic steroids, corticosteroids, and acidic drugs in horse urine by solid-phase extraction and liquid chromatography-mass spectrometry. *J Chromatogr A*: Epub 2006 May 2002.

Hoffmann Ed, Stroobant V (2001) *Mass spectrometry : principles and applications*. Wiley, Chichester ; New York.

Holbro T, Hynes NE (2004) ErbB receptors: directing key signaling networks throughout life. *Annu Rev Pharmacol Toxicol* 44: 195-217.

Holbrook MR, Slakey LL, Gross DJ (2000) Thermodynamic mixing of molecular states of the epidermal growth factor receptor modulates macroscopic ligand binding affinity. *Biochem J* 1: 99-108.

Hopfgartner G, Varesio E, Tschappat V, Grivet C, Bourgogne E, Leuthold LA (2004) Triple quadrupole linear ion trap mass spectrometer for the analysis of small molecules and macromolecules. *J Mass Spectrom* 39: 845-855.

Hsia DA, Mitra SK, Hauck CR, Streblov DN, Nelson JA, Ilic D, Huang S, Li E, Nemerow GR, Leng J, Spencer KS, Cheresch DA, Schlaepfer DD (2003) Differential regulation of cell motility and invasion by FAK. *J Cell Biol* 160: 753-767.

Huang C, Jacobson K, Schaller MD (2004) MAP kinases and cell migration. *J Cell Sci* 117: 4619-4628.

Hunt DF, Buko AM, Ballard JM, Shabanowitz J, Giordani AB (1981) Sequence analysis of polypeptides by collision activated dissociation on a triple quadrupole mass spectrometer. *Biomed Mass Spectrom* 8: 397-408.

Hunt DF, Yates JR, 3rd, Shabanowitz J, Winston S, Hauer CR (1986) Protein sequencing by tandem mass spectrometry. *Proc Natl Acad Sci U S A* 83: 6233-6237.

Irish JM, Hovland R, Krutzik PO, Perez OD, Bruserud O, Gjertsen BT, Nolan GP (2004) Single cell profiling of potentiated phospho-protein networks in cancer cells. *Cell* 118: 217-228.

Ito Y, Takeda T, Sakon M, Tsujimoto M, Higashiyama S, Noda K, Miyoshi E, Monden M, Matsuura N (2001) Expression and clinical significance of erb-B receptor family in hepatocellular carcinoma. *Br J Cancer* 84: 1377-1383.

Jain N, Zhang T, Kee WH, Li W, Cao X (1999) Protein kinase C delta associates with and phosphorylates Stat3 in an interleukin-6-dependent manner. *J Biol Chem* 274: 24392-24400.

Jaken S, Parker PJ (2000) Protein kinase C binding partners. *Bioessays* 22: 245-254.

Jawhari AU, Farthing MJ, Pignatelli M (1999) The E-cadherin/epidermal growth factor receptor interaction: a hypothesis of reciprocal and reversible control of intercellular adhesion and cell proliferation. *J Pathol* 187: 155-157.

Jia Z, Barbier L, Stuart H, Amraei M, Pelech S, Dennis JW, Metalnikov P, O'Donnell P, Nabi IR (2005) Tumor cell pseudopodial protrusions. Localized signaling domains coordinating cytoskeleton remodeling, cell adhesion, glycolysis, RNA translocation, and protein translation. *J Biol Chem*: Epub 2005 Jun 2028.

Jimenez de Asua L, Goin M (1992) Prostaglandin F2 alpha decreases the affinity of epidermal growth factor receptors in Swiss mouse 3T3 cells via protein kinase C activation. *FEBS Lett* 299: 235-238.

Joel PB, Smith J, Sturgill TW, Fisher TL, Blenis J, Lannigan DA (1998) pp90rsk1 regulates estrogen receptor-mediated transcription through phosphorylation of Ser-167. *Mol Cell Biol* 18: 1978-1984.

Jones FE, Stern DF (1999) Expression of dominant-negative ErbB2 in the mammary gland of transgenic mice reveals a role in lobuloalveolar development and lactation. *Oncogene* 18: 3481-3490.

Jones FE, Welte T, Fu XY, Stern DF (1999) ErbB4 signaling in the mammary gland is required for lobuloalveolar development and Stat5 activation during lactation. *J Cell Biol* 147: 77-88.

Karas M, Hillenkamp F (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 60: 2299-2301.

Karunakaran D, Tzahar E, Beerli RR, Chen X, Graus-Porta D, Ratzkin BJ, Seger R, Hynes NE, Yarden Y (1996) ErbB-2 is a common auxiliary subunit of NDF and EGF receptors: implications for breast cancer. *Embo J* 15: 254-264.

Kelleher NL, Zubarev RA, Bush K, Furie B, Furie BC, McLafferty FW, Walsh CT (1999) Localization of labile posttranslational modifications by electron capture dissociation: the case of gamma-carboxyglutamic acid. *Anal Chem* 71: 4250-4253.

Kim HK, Kim JW, Zilberstein A, Margolis B, Kim JG, Schlessinger J, Rhee SG (1991) PDGF stimulation of inositol phospholipid hydrolysis requires PLC-gamma 1 phosphorylation on tyrosine residues 783 and 1254. *Cell* 65: 435-441.

Kim JE, Tannenbaum SR, White FM (2005) Global phosphoproteome of HT-29 human colon adenocarcinoma cells. *J Proteome Res* 4: 1339-1346.

Kim JE, White FM (2006) Quantitative analysis of phosphotyrosine signaling networks triggered by CD3 and CD28 costimulation in Jurkat cells. *J Immunol* 176: 2833-2843.

Kim JY, Sun Q, Oglesbee M, Yoon SO (2003) The role of ErbB2 signaling in the onset of terminal differentiation of oligodendrocytes in vivo. *J Neurosci* 23: 5561-5571.

King CR, Kraus MH, Aaronson SA (1985) Amplification of a novel v-erbB-related gene in a human mammary carcinoma. *Science* 229: 974-976.

Kira M, Sano S, Takagi S, Yoshikawa K, Takeda J, Itami S (2002) STAT3 deficiency in keratinocytes leads to compromised cell migration through hyperphosphorylation of p130(cas). *J Biol Chem*: Epub 2002 Jan 2025.

Kirkpatrick DS, Gerber SA, Gygi SP (2005) The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods*: Epub 2005 Jan 2012.

Klapper LN, Glathe S, Vaisman N, Hynes NE, Andrews GC, Sela M, Yarden Y (1999) The ErbB-2/HER2 oncoprotein of human carcinomas may function solely as a shared coreceptor for multiple stroma-derived growth factors. *Proc Natl Acad Sci U S A* 96: 4995-5000.

Klein P, Mattoon D, Lemmon MA, Schlessinger J (2004) A structure-based model for ligand binding and dimerization of EGF receptors. *Proc Natl Acad Sci U S A*: Epub 2004 Jan 2019.

Knight ZA, Schilling B, Row RH, Kenski DM, Gibson BW, Shokat KM (2003) Phosphospecific proteolysis for mapping sites of protein phosphorylation. *Nat Biotechnol* 21: 1047-1054.

Koc H, Swenberg JA (2002) Applications of mass spectrometry for quantitation of DNA adducts. *J Chromatogr B Analyt Technol Biomed Life Sci* 778: 323-343.

Kratchmarova I, Blagoev B, Haack-Sorensen M, Kassem M, Mann M (2005) Mechanism of divergent growth factor effects in mesenchymal stem cell differentiation. *Science* 308: 1472-1477.

Kraus MH, Issing W, Miki T, Popescu NC, Aaronson SA (1989) Isolation and characterization of ERBB3, a third member of the ERBB/epidermal growth factor receptor family: evidence for overexpression in a subset of human mammary tumors. *Proc Natl Acad Sci U S A* 86: 9193-9197.

Krutzik PO, Irish JM, Nolan GP, Perez OD (2004) Analysis of protein phosphorylation and cellular signaling events by flow cytometry: techniques and clinical applications. *Clin Immunol* 110: 206-221.

Lacroix H, Iglehart JD, Skinner MA, Kraus MH (1989) Overexpression of erbB-2 or EGF receptor proteins present in early stage mammary carcinoma is detected simultaneously in matched primary tumors and regional metastases. *Oncogene* 4: 145-151.

Lawlor MA, Alessi DR (2001) PKB/Akt: a key mediator of cell proliferation, survival and insulin responses? *J Cell Sci* 114: 2903-2910.

Lax I, Bellot F, Howk R, Ullrich A, Givol D, Schlessinger J (1989) Functional analysis of the ligand binding site of EGF-receptor utilizing chimeric chicken/human receptor molecules. *Embo J* 8: 421-427.

Lee KF, Simon H, Chen H, Bates B, Hung MC, Hauser C (1995) Requirement for neuregulin receptor erbB2 in neural and cardiac development. *Nature* 378: 394-398.

Leitges M, Gimborn K, Elis W, Kalesnikoff J, Hughes MR, Krystal G, Huber M (2002) Protein kinase C-delta is a negative regulator of antigen-induced mast cell degranulation. *Mol Cell Biol* 22: 3970-3980.

Levi-Montalcini R, Cohen S (1960) Effects of the extract of the mouse submaxillary salivary glands on the sympathetic system of mammals. *Ann N Y Acad Sci* 85: 324-341.

Levkowitz G, Waterman H, Zamir E, Kam Z, Oved S, Langdon WY, Beguinot L, Geiger B, Yarden Y (1998) c-Cbl/Sli-1 regulates endocytic sorting and ubiquitination of the epidermal growth factor receptor. *Genes Dev* 12: 3663-3674.

Lewis GD, Lofgren JA, McMurtrey AE, Nuijens A, Fendly BM, Bauer KD, Sliwkowski MX (1996) Growth regulation of human breast and ovarian tumor

cells by heregulin: Evidence for the requirement of ErbB2 as a critical component in mediating heregulin responsiveness. *Cancer Res* 56: 1457-1465.

Li L, Backer J, Wong AS, Schwanke EL, Stewart BG, Pasdar M (2003) Bcl-2 expression decreases cadherin-mediated cell-cell adhesion. *J Cell Sci* 116: 3687-3700.

Liao H, Wu J, Kuhn E, Chin W, Chang B, Jones MD, O'Neil S, Clauser KR, Karl J, Hasler F, Roubenoff R, Zolg W, Guild BC (2004) Use of mass spectrometry to identify protein biomarkers of disease severity in the synovial fluid and serum of patients with rheumatoid arthritis. *Arthritis Rheum* 50: 3792-3803.

Lin LL, Wartmann M, Lin AY, Knopf JL, Seth A, Davis RJ (1993) cPLA2 is phosphorylated and activated by MAP kinase. *Cell* 72: 269-278.

Liu P, Rudick M, Anderson RG (2002) Multiple functions of caveolin-1. *J Biol Chem*: Epub 2002 Aug 2019.

Liu Y, Patricelli MP, Cravatt BF (1999) Activity-based protein profiling: the serine hydrolases. *Proc Natl Acad Sci U S A* 96: 14694-14699.

Lonardo F, Di Marco E, King CR, Pierce JH, Segatto O, Aaronson SA, Di Fiore PP (1990) The normal erbB-2 product is an atypical receptor-like tyrosine kinase with constitutive activity in the absence of ligand. *New Biol* 2: 992-1003.

Lorch JH, Klessner J, Park JK, Getsios S, Wu YL, Stack MS, Green KJ (2004) Epidermal growth factor receptor inhibition promotes desmosome assembly and strengthens intercellular adhesion in squamous cell carcinoma cells. *J Biol Chem* 279: 37191-37200.

Loyet KM, Stults JT, Arnott D (2005) Mass spectrometric contributions to the practice of phosphorylation site mapping through 2003: a literature review. *Mol Cell Proteomics*: Epub 2005 Jan 2007.

Lu Q, Mukhopadhyay NK, Griffin JD, Paredes M, Medina M, Kosik KS (2002) Brain armadillo protein delta-catenin interacts with Abl tyrosine kinase and modulates cellular morphogenesis in response to growth factors. *J Neurosci Res* 67: 618-624.

Lu Q, Paredes M, Medina M, Zhou J, Cavallo R, Peifer M, Orecchio L, Kosik KS (1999) delta-catenin, an adhesive junction-associated protein which promotes cell scattering. *J Cell Biol* 144: 519-532.

Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, Haluska FG, Louis DN, Christiani DC, Settleman J, Haber DA (2004) Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 350: 2129-2139.

Mackay K, Mochly-Rosen D (2001) Localization, anchoring, and functions of protein kinase C isozymes in the heart. *J Mol Cell Cardiol* 33: 1301-1307.

Mariner DJ, Davis MA, Reynolds AB (2004) EGFR signaling to p120-catenin through phosphorylation at Y228. *J Cell Sci* 117: 1339-1350.

Marmor MD, Skaria KB, Yarden Y (2004) Signal transduction and oncogenesis by ErbB/HER receptors. *Int J Radiat Oncol Biol Phys* 58: 903-913.

Martin-Fernandez M, Clarke DT, Tobin MJ, Jones SV, Jones GR (2002) Preformed oligomeric epidermal growth factor receptors undergo an ectodomain structure change during signaling. *Biophys J* 82: 2415-2427.

Mattoon D, Klein P, Lemmon MA, Lax I, Schlessinger J (2004) The tethered configuration of the EGF receptor extracellular domain exerts only a limited control of receptor function. *Proc Natl Acad Sci U S A*: Epub 2004 Jan 2019.

Maurer CA, Friess H, Kretschmann B, Zimmermann A, Stauffer A, Baer HU, Korc M, Buchler MW (1998) Increased expression of erbB3 in colorectal cancer is associated with concomitant increase in the level of erbB2. *Hum Pathol* 29: 771-777.

Mayawala K, Vlachos DG, Edwards JS (2005) Heterogeneities in EGF receptor density at the cell surface can lead to concave up scatchard plot of EGF binding. *FEBS Lett* 579: 3043-3047.

Mazumdar A, Adam L, Boyd D, Kumar R (2001) Heregulin regulation of urokinase plasminogen activator and its receptor: human breast epithelial cell invasion. *Cancer Res* 61: 400-405.

McLafferty FW, Horn DM, Breuker K, Ge Y, Lewis MA, Cerda B, Zubarev RA, Carpenter BK (2001) Electron capture dissociation of gaseous multiply charged ions by Fourier-transform ion cyclotron resonance. *J Am Soc Mass Spectrom* 12: 245-249.

Meier R, Hemmings BA (1999) Regulation of protein kinase B. *J Recept Signal Transduct Res* 19: 121-128.

Meiners S, Brinkmann V, Naundorf H, Birchmeier W (1998) Role of morphogenetic factors in metastasis of mammary carcinoma cells. *Oncogene* 16: 9-20.

Melanson JE, Chisholm KA, Pinto DM (2006) Targeted comparative proteomics by liquid chromatography/matrix-assisted laser desorption/ionization triple-quadrupole mass spectrometry. *Rapid Commun Mass Spectrom* 20: 904-910.

Menard S, Casalini P, Campiglio M, Pupa SM, Tagliabue E (2004) Role of HER2/neu in tumor progression and therapy. *Cell Mol Life Sci* 61: 2965-2978.

Mendelsohn J, Baselga J (2003) Status of epidermal growth factor receptor antagonists in the biology and treatment of cancer. *J Clin Oncol* 21: 2787-2799.

Mendelsohn J, Baselga J (2006) Epidermal growth factor receptor targeting in cancer. *Semin Oncol* 33: 369-385.

Meng F, Forbes AJ, Miller LM, Kelleher NL (2005) Detection and localization of protein modifications by high resolution tandem mass spectrometry. *Mass Spectrom Rev* 24: 126-134.

Meyer D, Yamaai T, Garratt A, Riethmacher-Sonnenberg E, Kane D, Theill LE, Birchmeier C (1997) Isoform-specific expression and function of neuregulin. *Development* 124: 3575-3586.

Miettinen PJ, Berger JE, Meneses J, Phung Y, Pedersen RA, Werb Z, Derynck R (1995) Epithelial immaturity and multiorgan failure in mice lacking epidermal growth factor receptor. *Nature* 376: 337-341.

Mirgorodskaya E, Wanker E, Otto A, Lehrach H, Gobom J (2005) Method for qualitative comparisons of protein mixtures based on enzyme-catalyzed stable-isotope incorporation. *J Proteome Res* 4: 2109-2116.

Mirgorodskaya OA, Kozmin YP, Titov MI, Korner R, Sonksen CP, Roepstorff P (2000) Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using (18)O-labeled internal standards. *Rapid Commun Mass Spectrom* 14: 1226-1232.

Moore RE, Licklider L, Schumann D, Lee TD (1998) A microscale electrospray interface incorporating a monolithic, poly(styrene-divinylbenzene) support for on-line liquid chromatography/tandem mass spectrometry analysis of peptides and proteins. *Anal Chem* 70: 4879-4884.

Moriki T, Maruyama H, Maruyama IN (2001) Activation of preformed EGF receptor dimers by ligand-induced rotation of the transmembrane domain. *J Mol Biol* 311: 1011-1026.

Moro L, Dolce L, Cabodi S, Bergatto E, Erba EB, Smeriglio M, Turco E, Retta SF, Giuffrida MG, Venturino M, Godovac-Zimmermann J, Conti A, Schaefer E, Beguinot L, Tacchetti C, Gaggini P, Silengo L, Tarone G, Defilippi P (2002) Integrin-induced epidermal growth factor (EGF) receptor activation requires c-Src and p130Cas and leads to phosphorylation of specific EGF receptor tyrosines. *J Biol Chem* 277: 9405-9414.

Morris JK, Lin W, Hauser C, Marchuk Y, Getman D, Lee KF (1999) Rescue of the cardiac defect in ErbB2 mutant mice reveals essential roles of ErbB2 in peripheral nervous system development. *Neuron* 23: 273-283.

Morrison P, Takishima K, Rosner MR (1993) Role of threonine residues in regulation of the epidermal growth factor receptor by protein kinase C and mitogen-activated protein kinase. *J Biol Chem* 268: 15536-15543.

Moscattello DK, Holgado-Madruga M, Godwin AK, Ramirez G, Gunn G, Zoltick PW, Biegel JA, Hayes RL, Wong AJ (1995) Frequent expression of a mutant epidermal growth factor receptor in multiple human tumors. *Cancer Res* 55: 5536-5539.

Moser K, White FM (2006) Phosphoproteomic analysis of rat liver by high capacity IMAC and LC-MS/MS. *J Proteome Res* 5: 98-104.

Motoyama A, Venable JD, Ruse CI, Yates JR, 3rd (2006) Automated Ultra-High-Pressure Multidimensional Protein Identification Technology (UHP-MudPIT) for Improved Peptide Identification of Proteomic Samples. *Anal Chem* 78: 5109-5118.

Munchbach M, Quadroni M, Miotto G, James P (2000) Quantitation and facilitated de novo sequencing of proteins by isotopic N-terminal labeling of peptides with a fragmentation-directing moiety. *Anal Chem* 72: 4047-4057.

Navolanic PM, Steelman LS, McCubrey JA (2003) EGFR family signaling and its association with breast cancer development and resistance to chemotherapy (Review). *Int J Oncol* 22: 237-252.

Neve RM, Lane HA, Hynes NE (2001) The role of overexpressed HER2 in transformation. *Ann Oncol* 12 Suppl 1: S9-13.

Niemann C, Brinkmann V, Spitzer E, Hartmann G, Sachs M, Naundorf H, Birchmeier W (1998) Reconstitution of mammary gland development in vitro: requirement of c-met and c-erbB2 signaling for branching and alveolar morphogenesis. *J Cell Biol* 143: 533-545.

Nollet F, Berx G, van Roy F (1999) The role of the E-cadherin/catenin adhesion complex in the development and progression of cancer. *Mol Cell Biol Res Commun* 2: 77-85.

Normanno N, De Luca A, Bianco C, Strizzi L, Mancino M, Maiello MR, Carotenuto A, De Feo G, Caponigro F, Salomon DS (2006) Epidermal growth factor receptor (EGFR) signaling in cancer. *Gene*: Epub 2005 Dec 2027.

Oda Y, Nagasu T, Chait BT (2001) Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat Biotechnol* 19: 379-382.

Ogiso H, Ishitani R, Nureki O, Fukai S, Yamanaka M, Kim JH, Saito K, Sakamoto A, Inoue M, Shirouzu M, Yokoyama S (2002) Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell* 110: 775-787.

Okabayashi Y, Kido Y, Okutani T, Sugimoto Y, Sakaguchi K, Kasuga M (1994) Tyrosines 1148 and 1173 of activated human epidermal growth factor receptors are binding sites of Shc in intact cells. *J Biol Chem* 269: 18674-18678.

Olapade-Olaopa EO, Moscatello DK, MacKay EH, Sandhu DP, Terry TR, Wong AJ, Habib FK (2004) Alterations in the expression of androgen receptor, wild type-epidermal growth factor receptor and a mutant epidermal growth factor receptor in human prostate cancer. *Afr J Med Med Sci* 33: 245-253.

Oliva JL, Griner EM, Kazanietz MG (2005) PKC isozymes and diacylglycerol-regulated proteins as effectors of growth factor receptors. *Growth Factors* 23: 245-252.

Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* 1: 376-386.

Ozcelik C, Erdmann B, Pilz B, Wettschureck N, Britsch S, Hubner N, Chien KR, Birchmeier C, Garratt AN (2002) Conditional mutation of the ErbB2 (HER2) receptor in cardiomyocytes leads to dilated cardiomyopathy. *Proc Natl Acad Sci U S A*: Epub 2002 Jun 2018.

Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, Fujii Y, Eck MJ, Sellers WR, Johnson BE, Meyerson M (2004) EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 304: 1497-1500.

Pandey A, Andersen JS, Mann M (2000a) Use of mass spectrometry to study signaling pathways. *Sci STKE* 2000: PL1.

Pandey A, Blagoev B, Kratchmarova I, Fernandez M, Nielsen M, Kristiansen TZ, Ohara O, Podtelejnikov AV, Roche S, Lodish HF, Mann M (2002) Cloning of a novel phosphotyrosine binding domain containing molecule, Odin, involved in signaling by receptor tyrosine kinases. *Oncogene* 21: 8029-8036.

Pandey A, Podtelejnikov AV, Blagoev B, Bustelo XR, Mann M, Lodish HF (2000b) Analysis of receptor signaling pathways by mass spectrometry: identification of vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors. *Proc Natl Acad Sci U S A* 97: 179-184.

Patterson RL, van Rossum DB, Nikolaidis N, Gill DL, Snyder SH (2005) Phospholipase C-gamma: diverse roles in receptor-mediated calcium signaling. *Trends Biochem Sci*: Epub 2005 Nov 2002.

Pedersen MW, Meltorn M, Damstrup L, Poulsen HS (2001) The type III epidermal growth factor receptor mutation. Biological significance and potential target for anti-cancer therapy. *Ann Oncol* 12: 745-760.

Pelkmans L, Helenius A (2002) Endocytosis via caveolae. *Traffic* 3: 311-320.

Pende M, Fisher TL, Simpson PB, Russell JT, Blenis J, Gallo V (1997) Neurotransmitter- and growth factor-induced cAMP response element binding protein phosphorylation in glial cell progenitors: role of calcium ions, protein kinase C, and mitogen-activated protein kinase/ribosomal S6 kinase pathway. *J Neurosci* 17: 1291-1301.

Perez R, Pascual M, Macias A, Lage A (1984) Epidermal growth factor receptors in human breast cancer. *Breast Cancer Res Treat* 4: 189-193.

Piepkorn M, Pittelkow MR, Cook PW (1998) Autocrine regulation of keratinocytes: the emerging role of heparin-binding, epidermal growth factor-related growth factors. *J Invest Dermatol* 111: 715-721.

Pigini D, Cialdella AM, Faranda P, Tranfo G (2006) Comparison between external and internal standard calibration in the validation of an analytical method for 1-hydroxypyrene in human urine by high-performance liquid chromatography/tandem mass spectrometry. *Rapid Commun Mass Spectrom* 20: 1013-1018.

Pinkas-Kramarski R, Soussan L, Waterman H, Levkowitz G, Alroy I, Klapper L, Lavi S, Seger R, Ratzkin BJ, Sela M, Yarden Y (1996) Diversification of Neu differentiation factor and epidermal growth factor signaling by combinatorial receptor interactions. *Embo J* 15: 2452-2467.

Playford MP, Schaller MD (2004) The interplay between Src and integrins in normal and tumor biology. *Oncogene* 23: 7928-7946.

Plowman GD, Culouscou JM, Whitney GS, Green JM, Carlton GW, Foy L, Neubauer MG, Shoyab M (1993) Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal growth factor receptor family. *Proc Natl Acad Sci U S A* 90: 1746-1750.

Prenzel N, Fischer OM, Streit S, Hart S, Ullrich A (2001) The epidermal growth factor receptor family as a central element for cellular signal transduction and diversification. *Endocr Relat Cancer* 8: 11-31.

Rasheed BK, Wiltshire RN, Bigner SH, Bigner DD (1999) Molecular pathogenesis of malignant gliomas. *Curr Opin Oncol* 11: 162-167.

Raught B, Gingras AC (1999) eIF4E activity is regulated at multiple levels. *Int J Biochem Cell Biol* 31: 43-57.

Razani B, Schlegel A, Liu J, Lisanti MP (2001) Caveolin-1, a putative tumour suppressor gene. *Biochem Soc Trans* 29: 494-499.

Reinders J, Sickmann A (2005) State-of-the-art in phosphoproteomics. *Proteomics* 5: 4052-4061.

Reynolds AB, Herbert L, Cleveland JL, Berg ST, Gaut JR (1992) p120, a novel substrate of protein tyrosine kinase receptors and of p60v-src, is related to cadherin-binding factors beta-catenin, plakoglobin and armadillo. *Oncogene* 7: 2439-2445.

Reynolds KJ, Yao X, Fenselau C (2002) Proteolytic ¹⁸O labeling for comparative proteomics: evaluation of endoprotease Glu-C as the catalytic agent. *J Proteome Res* 1: 27-33.

Riethmacher D, Sonnenberg-Riethmacher E, Brinkmann V, Yamaai T, Lewin GR, Birchmeier C (1997) Severe neuropathies in mice with targeted mutations in the ErbB3 receptor. *Nature* 389: 725-730.

Rivera VM, Miranti CK, Misra RP, Ginty DD, Chen RH, Blenis J, Greenberg ME (1993) A growth factor-induced kinase phosphorylates the serum response factor at a site that regulates its DNA-binding activity. *Mol Cell Biol* 13: 6260-6273.

Rodrigues GA, Falasca M, Zhang Z, Ong SH, Schlessinger J (2000) A novel positive feedback loop mediated by the docking protein Gab1 and phosphatidylinositol 3-kinase in epidermal growth factor receptor signaling. *Mol Cell Biol* 20: 1448-1459.

Roepstorff P, Fohlman J (1984) Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* 11: 601.

Rogers SJ, Harrington KJ, Rhys-Evans P, P OC, Eccles SA (2005) Biological significance of c-erbB family oncogenes in head and neck cancer. *Cancer Metastasis Rev* 24: 47-69.

Roskoski R, Jr. (2004) Src protein-tyrosine kinase structure and regulation. *Biochem Biophys Res Commun* 324: 1155-1164.

Ross JS, Fletcher JA, Bloom KJ, Linette GP, Stec J, Symmans WF, Puztai L, Hortobagyi GN (2004a) Targeted therapy in breast cancer: the HER-2/neu gene and protein. *Mol Cell Proteomics* 3: 379-398.

Ross JS, Fletcher JA, Linette GP, Stec J, Clark E, Ayers M, Symmans WF, Puztai L, Bloom KJ (2003) The Her-2/neu gene and protein in breast cancer 2003: biomarker and target of therapy. *Oncologist* 8: 307-325.

Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlet-Jones M, He F, Jacobson A, Pappin DJ (2004b) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* 3: 1154-1169.

Rush J, Moritz A, Lee KA, Guo A, Goss VL, Spek EJ, Zhang H, Zha XM, Polakiewicz RD, Comb MJ (2004) Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat Biotechnol*.

Rusnak F, Zhou J, Hathaway GM (2002) Identification of phosphorylated and glycosylated sites in peptides by chemically targeted proteolysis. *Journal of Biomolecular Techniques* 13: 228-237.

Sako Y, Minoghchi S, Yanagida T (2000) Single-molecule imaging of EGFR signalling on the surface of living cells. *Nat Cell Biol* 2: 168-172.

Salomon AR, Ficarro SB, Brill LM, Brinker A, Phung QT, Ericson C, Sauer K, Brock A, Horn DM, Schultz PG, Peters EC (2003) Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry. *Proc Natl Acad Sci U S A* 100: 443-448.

Salomon DS, Brandt R, Ciardiello F, Normanno N (1995) Epidermal growth factor-related peptides and their receptors in human malignancies. *Crit Rev Oncol Hematol* 19: 183-232.

Sarbassov DD, Guertin DA, Ali SM, Sabatini DM (2005) Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. *Science* 307: 1098-1101.

Sato JD, Kawamoto T, Le AD, Mendelsohn J, Polikoff J, Sato GH (1983) Biological effects in vitro of monoclonal antibodies to human epidermal growth factor receptors. *Mol Biol Med* 1: 511-529.

Scagliotti GV, Selvaggi G, Novello S, Hirsch FR (2004) The biology of epidermal growth factor receptor in lung cancer. *Clin Cancer Res* 10: 4227s-4232s.

Schafer B, Gschwind A, Ullrich A (2004) Multiple G-protein-coupled receptor signals converge on the epidermal growth factor receptor to promote migration and invasion. *Oncogene* 23: 991-999.

Schaller MD, Hildebrand JD, Shannon JD, Fox JW, Vines RR, Parsons JT (1994) Autophosphorylation of the focal adhesion kinase, pp125FAK, directs SH2-dependent binding of pp60src. *Mol Cell Biol* 14: 1680-1688.

Schaller MD, Parsons JT (1995) pp125FAK-dependent tyrosine phosphorylation of paxillin creates a high-affinity binding site for Crk. *Mol Cell Biol* 15: 2635-2645.

Schlegel A, Pestell RG, Lisanti MP (2000) Caveolins in cholesterol trafficking and signal transduction: implications for human disease. *Front Biosci* 1: D929-937.

Schlessinger J (1988) The epidermal growth factor receptor as a multifunctional allosteric protein. *Biochemistry* 27: 3119-3123.

Schlosser A, Pipkorn R, Bossemeyer D, Lehmann WD (2001) Analysis of protein phosphorylation by a combination of elastase digestion and neutral loss tandem mass spectrometry. *Anal Chem* 73: 170-176.

Schmelzle K, Kane S, Gridley S, Lienhard GE, White FM (2006) Temporal dynamics of tyrosine phosphorylation in insulin signaling. *Diabetes* 55: 2171-2179.

Schmelzle K, White FM (2006) Phosphoproteomic approaches to elucidate cellular signaling networks. *Curr Opin Biotechnol*: Epub 2006 Jun 2027.

Schoeberl B, Eichler-Jonsson C, Gilles ED, Muller G (2002) Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat Biotechnol* 20: 370-375.

Schroeder JA, Lee DC (1998) Dynamic expression and activation of ERBB receptors in the developing mouse mammary gland. *Cell Growth Differ* 9: 451-464.

Sebastian J, Richards RG, Walker MP, Wiesen JF, Werb Z, Derynck R, Hom YK, Cunha GR, DiAugustine RP (1998) Activation and function of the epidermal growth factor receptor and erbB-2 during mammary gland morphogenesis. *Cell Growth Differ* 9: 777-785.

Semba K, Kamata N, Toyoshima K, Yamamoto T (1985) A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epidermal growth factor-receptor gene and is amplified in a human salivary gland adenocarcinoma. *Proc Natl Acad Sci U S A* 82: 6497-6501.

Shinojima N, Tada K, Shiraishi S, Kamiryo T, Kochi M, Nakamura H, Makino K, Saya H, Hirano H, Kuratsu J, Oka K, Ishimaru Y, Ushio Y (2003) Prognostic value of epidermal growth factor receptor in patients with glioblastoma multiforme. *Cancer Res* 63: 6962-6970.

Sibilia M, Steinbach JP, Stingl L, Aguzzi A, Wagner EF (1998) A strain-independent postnatal neurodegeneration in mice lacking the EGF receptor. *Embo J* 17: 719-731.

Sibilia M, Wagner EF (1995) Strain-dependent epithelial defects in mice lacking the EGF receptor. *Science* 269: 234-238.

Sierke SL, Cheng K, Kim HH, Koland JG (1997) Biochemical characterization of the protein tyrosine kinase homology domain of the ErbB3 (HER3) receptor protein. *Biochem J* 322: 757-763.

Silva CM (2004) Role of STATs as downstream signal transducers in Src family kinase-mediated tumorigenesis. *Oncogene* 23: 8017-8023.

Silver DL, Naora H, Liu J, Cheng W, Montell DJ (2004) Activated signal transducer and activator of transcription (STAT) 3: localization in focal adhesions and function in ovarian cancer cell motility. *Cancer Res* 64: 3550-3558.

Simpson BJ, Weatherill J, Miller EP, Lessells AM, Langdon SP, Miller WR (1995) c-erbB-3 protein expression in ovarian tumours. *Br J Cancer* 71: 758-762.

Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming T, Eiermann W, Wolter J, Pegram M, Baselga J, Norton L (2001) Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 344: 783-792.

Sleno L, Volmer DA (2004) Ion activation methods for tandem mass spectrometry. *J Mass Spectrom* 39: 1091-1112.

Soloaga A, Thomson S, Wiggin GR, Rampersaud N, Dyson MH, Hazzalin CA, Mahadevan LC, Arthur JS (2003) MSK2 and MSK1 mediate the mitogen- and stress-induced phosphorylation of histone H3 and HMG-14. *Embo J* 22: 2788-2797.

Soltoff SP, Carraway KL, 3rd, Prigent SA, Gullick WG, Cantley LC (1994) ErbB3 is involved in activation of phosphatidylinositol 3-kinase by epidermal growth factor. *Mol Cell Biol* 14: 3550-3558.

Stamos J, Sliwkowski MX, Eigenbrot C (2002) Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J Biol Chem*: Epub 2002 Aug 2023.

Steen H, Fernandez M, Ghaffari S, Pandey A, Mann M (2003) Phosphotyrosine mapping in Bcr/Abl oncoprotein using phosphotyrosine-specific immonium ion scanning. *Mol Cell Proteomics*: Epub 2003 Feb 2025.

Steen H, Kuster B, Fernandez M, Pandey A, Mann M (2001) Detection of tyrosine phosphorylated peptides by precursor ion scanning quadrupole TOF mass spectrometry in positive ion mode. *Anal Chem* 73: 1440-1448.

Steen H, Kuster B, Fernandez M, Pandey A, Mann M (2002) Tyrosine phosphorylation mapping of the epidermal growth factor receptor signaling pathway. *J Biol Chem* 277: 1031-1039.

Steinberg SF (2004) Distinctive activation mechanisms and functions for protein kinase Cdelta. *Biochem J* 384: 449-459.

Stern DF (2003) ErbBs in mammary development. *Exp Cell Res* 284: 89-98.

Stove C, Bracke M (2004) Roles for neuregulins in human cancer. *Clin Exp Metastasis* 21: 665-684.

Stults JT, Arnott D (2005) Proteomics. *Methods Enzymol* 402: 245-289.

Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A* 101: 9528-9533.

Takeichi M (1995) Morphogenetic roles of classic cadherins. *Curr Opin Cell Biol* 7: 619-627.

Tang CK, Perez C, Grunt T, Waibel C, Cho C, Lupu R (1996) Involvement of heregulin-beta2 in the acquisition of the hormone-independent phenotype of breast cancer cells. *Cancer Res* 56: 3350-3358.

Thevis M, Opfermann G, Schanzer W (2001) High speed determination of beta-receptor blocking agents in human urine by liquid chromatography/tandem mass spectrometry. *Biomed Chromatogr* 15: 393-402.

Thomson JJ (1913) *Rays of Positive Electricity and Their Application to Chemical Analysis*. Longmans Green, London.

Thor AD, Liu S, Edgerton S, Moore D, 2nd, Kasowitz KM, Benz CC, Stern DF, DiGiovanna MP (2000) Activation (tyrosine phosphorylation) of ErbB-2 (HER-2/neu): a study of incidence and correlation with outcome in breast cancer. *J Clin Oncol* 18: 3230-3239.

Threadgill DW, Dlugosz AA, Hansen LA, Tennenbaum T, Lichti U, Yee D, LaMantia C, Mourton T, Herrup K, Harris RC, et al. (1995) Targeted disruption of mouse EGF receptor: effect of genetic background on mutant phenotype. *Science* 269: 230-234.

Tidcombe H, Jackson-Fisher A, Mathers K, Stern DF, Gassmann M, Golding JP (2003) Neural and mammary gland defects in ErbB4 knockout mice genetically rescued from embryonic lethality. *Proc Natl Acad Sci U S A*: Epub 2003 Jun 2024.

Torrisi MR, Lotti LV, Belleudi F, Gradini R, Salcini AE, Confalonieri S, Pelicci PG, Di Fiore PP (1999) Eps15 is recruited to the plasma membrane upon epidermal growth factor receptor activation and localizes to components of the endocytic pathway during receptor internalization. *Mol Biol Cell* 10: 417-434.

Tsai MS, Hornby AE, Lakins J, Lupu R (2000) Expression and function of CYR61, an angiogenic factor, in breast cancer cell lines and tumor biopsies. *Cancer Res* 60: 5603-5607.

Tsai MS, Shamon-Taylor LA, Mehmi I, Tang CK, Lupu R (2003) Blockage of heregulin expression inhibits tumorigenicity and metastasis of breast cancer. *Oncogene* 22: 761-768.

Uddin S, Sassano A, Deb DK, Verma A, Majchrzak B, Rahman A, Malik AB, Fish EN, Plataniias LC (2002) Protein kinase C-delta (PKC-delta) is activated by type I interferons and mediates phosphorylation of Stat1 on serine 727. *J Biol Chem*: Epub 2002 Feb 2011.

Vadlamudi RK, Sahin AA, Adam L, Wang RA, Kumar R (2003) Heregulin and HER2 signaling selectively activates c-Src phosphorylation at tyrosine 215. *FEBS Lett* 543: 76-80.

Vuori K, Hirai H, Aizawa S, Ruoslahti E (1996) Introduction of p130cas signaling complex formation upon integrin-mediated cell adhesion: a role for Src family kinases. *Mol Cell Biol* 16: 2606-2613.

Walker F, Orchard SG, Jorissen RN, Hall NE, Zhang HH, Hoyne PA, Adams TE, Johns TG, Ward C, Garrett TP, Zhu HJ, Nerrie M, Scott AM, Nice EC, Burgess AW (2004) CR1/CR2 interactions modulate the functions of the cell surface epidermal growth factor receptor. *J Biol Chem*: Epub 2004 Mar 2011.

Wallasch C, Weiss FU, Niederfellner G, Jallal B, Issing W, Ullrich A (1995) Heregulin-dependent regulation of HER2/neu oncogenic signaling by heterodimerization with HER3. *Embo J* 14: 4267-4275.

Wang HG, Pathan N, Ethell IM, Krajewski S, Yamaguchi Y, Shibasaki F, McKeon F, Bobo T, Franke TF, Reed JC (1999) Ca²⁺-induced apoptosis through calcineurin dephosphorylation of BAD. *Science* 284: 339-343.

Wells A, Grandis JR (2003) Phospholipase C-gamma1 in tumor progression. *Clin Exp Metastasis* 20: 285-290.

Wells A, Marti U (2002) Signalling shortcuts: cell-surface receptors in the nucleus? *Nat Rev Mol Cell Biol* 3: 697-702.

Wells A, Ware MF, Allen FD, Lauffenburger DA (1999) Shaping up for shipping out: PLCgamma signaling of morphology changes in EGF-stimulated fibroblast migration. *Cell Motil Cytoskeleton* 44: 227-233.

Wienkoop S, Weckwerth W (2006) Relative and absolute quantitative shotgun proteomics: targeting low-abundance proteins in *Arabidopsis thaliana*. *J Exp Bot*: Epub 2006 Mar 2030.

Wiggin GR, Soloaga A, Foster JM, Murray-Tait V, Cohen P, Arthur JS (2002) MSK1 and MSK2 are required for the mitogen- and stress-induced phosphorylation of CREB and ATF1 in fibroblasts. *Mol Cell Biol* 22: 2871-2881.

Wofsy C, Goldstein B, Lund K, Wiley HS (1992) Implications of epidermal growth factor (EGF) induced egf receptor aggregation. *Biophys J* 63: 98-110.

Wolpowitz D, Mason TB, Dietrich P, Mendelsohn M, Talmage DA, Role LW (2000) Cysteine-rich domain isoforms of the neuregulin-1 gene are required for maintenance of peripheral synapses. *Neuron* 25: 79-91.

Wong AJ, Ruppert JM, Bigner SH, Grzeschik CH, Humphrey PA, Bigner DS, Vogelstein B (1992) Structural alterations of the epidermal growth factor receptor gene in human gliomas. *Proc Natl Acad Sci U S A* 89: 2965-2969.

Worthylake R, Opresko LK, Wiley HS (1999) ErbB-2 amplification inhibits down-regulation and induces constitutive activation of both ErbB-2 and epidermal growth factor receptors. *J Biol Chem* 274: 8865-8874.

Wysocki VH, Resing KA, Zhang Q, Cheng G (2005) Mass spectrometry of peptides and proteins. *Methods*: Epub 2005 Jan 2020.

Xia W, Lau YK, Zhang HZ, Xiao FY, Johnston DA, Liu AR, Li L, Katz RL, Hung MC (1999) Combination of EGFR, HER-2/neu, and HER-3 is a stronger predictor for the outcome of oral squamous cell carcinoma than any individual family members. *Clin Cancer Res* 5: 4164-4174.

Xu YH, Ishii S, Clark AJ, Sullivan M, Wilson RK, Ma DP, Roe BA, Merlino GT, Pastan I (1984) Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs overproduced in A431 carcinoma cells. *Nature* 309: 806-810.

Yao J, Xiong S, Klos K, Nguyen N, Grijalva R, Li P, Yu D (2001a) Multiple signaling pathways involved in activation of matrix metalloproteinase-9 (MMP-9) by heregulin-beta1 in human breast cancer cells. *Oncogene* 20: 8066-8074.

Yao X, Freas A, Ramirez J, Demirev PA, Fenselau C (2001b) Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal Chem* 73: 2836-2842.

Yap AS (1998) The morphogenetic role of cadherin cell adhesion molecules in human cancer: a thematic review. *Cancer Invest* 16: 252-261.

Yarden Y, Schlessinger J (1987) Epidermal growth factor induces rapid, reversible aggregation of the purified epidermal growth factor receptor. *Biochemistry* 26: 1443-1451.

Yarden Y, Sliwkowski MX (2001) Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol* 2: 127-137.

Yokota J, Yamamoto T, Toyoshima K, Terada M, Sugimura T, Battifora H, Cline MJ (1986) Amplification of c-erbB-2 oncogene in human adenocarcinomas in vivo. *Lancet* 1: 765-767.

Yu X, Sharma KD, Takahashi T, Iwamoto R, Mekada E (2002) Ligand-independent dimer formation of epidermal growth factor receptor (EGFR) is a step separable from ligand-induced EGFR signaling. *Mol Biol Cell* 13: 2547-2557.

Zappacosta F, Collingwood TS, Huddleston MJ, Annan RS (2006) A quantitative results-driven approach to analyzing multi-site protein phosphorylation: The phosphate-dependent phosphorylation profile of the transcription factor pho4. *Mol Cell Proteomics* 9: 9.

Zeillinger R, Kury F, Czerwenka K, Kubista E, Sliutz G, Knogler W, Huber J, Zielinski C, Reiner G, Jakesz R, et al. (1989) HER-2 amplification, steroid receptors and epidermal growth factor receptor in primary breast cancer. *Oncogene* 4: 109-114.

Zenobi R, Knochenmuss R (1998) Ion formation in MALDI mass spectrometry. *Mass Spectrometry Reviews* 17: 337-366.

Zhang H, Li XJ, Martin DB, Aebersold R (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol*: Epub 2003 May 2018.

Zhang H, Yi EC, Li XJ, Mallick P, Kelly-Spratt KS, Masselon CD, Camp DG, 2nd, Smith RD, Kemp CJ, Aebersold R (2005a) High throughput quantitative analysis of serum proteins using glycopeptide capture and liquid chromatography mass spectrometry. *Mol Cell Proteomics*: Epub 2004 Dec 2017.

Zhang Y, Wolf-Yadlin A, Ross PL, Pappin DJ, Rush J, Lauffenburger DA, White FM (2005b) Time-resolved Mass Spectrometry of Tyrosine Phosphorylation Sites in the Epidermal Growth Factor Receptor Signaling Network Reveals Dynamic Modules. *Mol Cell Proteomics* 4: 1240-1250.

Zhou H, Ranish JA, Watts JD, Aebersold R (2002) Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. *Nat Biotechnol* 20: 512-515.

Zhou H, Watts JD, Aebersold R (2001) A systematic approach to the analysis of protein phosphorylation. *Nat Biotechnol* 19: 375-378.

Zubarev RA (2004) Electron-capture dissociation tandem mass spectrometry. *Curr Opin Biotechnol* 15: 12-16.

**II Antibodies against Domains II and IV of EGFR Inhibit
Receptor Phosphorylation and Decrease High-Affinity
Binding by Blocking Preformed Dimers**

II.1 SUMMARY

We have generated two rabbit polyclonal antibodies that bind either EGFR's domain II dimerization arm or a domain IV putative dimerization loop. We have studied the effect of these antibodies on EGF-EGFR binding, EGFR dimerization and EGFR mediated phosphorylation. We found that both antibodies abrogate the high affinity component of EGF-EGFR binding, but do not compete with ligand binding. They also block EGFR homo- and hetero-dimerization and inhibit EGFR mediated phosphorylation of itself and HER2.

Refined versions of these antibodies as humanized monoclonal antibodies hold great potential for cancer therapy, especially in tumors driven by EGFR autocrine loops or constitutively active mutants where ligand-competitive antibodies efficiency might be low.

Based on our experimental results Ginger Chao developed an equilibrium model for receptor dimerization and ligand binding. Details on the model will be part of Ginger's thesis and will also appear in a future publication co-authored by both of us (Wolf-Yadlin et al, work in progress).

II.2 MATERIALS AND METHODS

II.2.1 Peptides and antibodies

DII and DIV peptides were synthesized, conjugated to KLH, and used to raise polyclonal antibodies rabbits by Cell Essentials (Boston, MA). The rabbit antisera were purified using ammonium sulfate precipitation followed by dialysis. Monoclonal antibody 225 was secreted from hybridoma cells (American Type Culture Collection, ATCC) and purified using ammonium sulfate precipitation followed by protein G purification (Sigma). Anti-EGFR and anti-HER2 rabbit polyclonal antibodies for Western blots were purchased from Cell Signaling Technologies and used at 1:1000. Anti-EGFR pY1173 and anti-HER2 pY1248 antibodies were purchased from Santa Cruz Biotechnology and used at 1:100. Horseradish peroxidase-conjugated donkey anti-rabbit antibody was purchased from Amersham Biosciences. Phycoerythrin (PE)-labeled goat anti-rabbit antibody and Alexa488-labeled EGF were purchased from Molecular Probes.

II.2.2 Cell Culture

184A1 human mammary epithelial cells (Stampfer & Bartley, 1985) (parental HMECs) were a generous gift from Martha Stampfer (Lawrence Berkeley Laboratory, Berkeley, CA) and were maintained in DFCI-1 medium

supplemented with 12.5 ng/ml EGF (Peprotech) (Band & Sager, 1989). 184A1 HMECs over-expressing HER2 (HMECs 24H) were a kind gift from Steven Wiley (Pacific Northwest National Laboratory, Richland, WA) and were maintained in DFCI-1 medium supplemented with 12.5 ng/ml EGF and 150 μ g/ml G418 (Hendriks et al., 2003a).

II.2.3 Cell Treatment

Cells were grown to 80% confluence in 6-well or 10 cm plates. The cells were washed with phosphate-buffered saline (PBS) and serum starved overnight in serum-free media (DFCI-1 without sodium bicarbonate, EGF, bovine pituitary extract, and fetal bovine serum). The cells were then washed with PBS and incubated for 2 hour at 37°C in serum-free media containing various concentrations of the anti-DII or anti-DIV antibody. Following, the cells were again washed with PBS and incubated for 5 minutes at 37°C in serum-free media containing 25 nM EGF. Finally the cells were once more washed with PBS and lysed with lysis buffer as described in (Janes et al., 2003).

II.2.4 Protein Quantification

Protein concentration of cell lysates was determined using the micro bicinchoninic acid assay (micro-BSA – Pierce) according to the manufacturer's protocol.

II.2.5 Quantitative Western Blot Analysis

40 µg of cell lysates/well were mixed with 4x sample buffer (250 mM Tris-HCl pH 6.8, 8% SDS, 40% glycerol, 0.04% bromophenol blue, and 400 mM DTT) and boiled for 5 minutes. Cell lysates were then separated by SDS-PAGE and transferred onto PDVF membranes. Membranes were blocked with 5% non-fat milk powder in TBST (20 mM Tris-HCl pH 7.5, 137 mM NaCl, 0.1% Tween-20) for 1 hour at room temperature and incubated overnight at 4°C in primary antibody, washed 3 times for 5 minutes in TBST and incubated for 1 hour at room temperature in secondary antibody (1:5000 horseradish peroxidase-conjugated donkey anti-rabbit in TBS-T, 5% non-fat milk powder) washed again 3 times for 5 minutes with TBS-T. Blots were developed with ECL Advance western blotting detection kit (Amersham), scanned on a Kodak Image Station 1000 and quantified using Kodak 1D 3.5 software. Relative phosphorylation intensities were calculated by dividing the phosphorylation sites intensities by the total receptor intensity measurement and normalizing by the positive control ratio.

II.2.6 Flow Cytometry

Cells were grown and serum starved as above. The cells were then washed, detached using a cell scraper, spun down at 2500 rpm, and washed 3 times in PBS. The cells were incubated for 2 hours at room temperature with 200 μ l of 200 μ g/ml anti-DII or anti-DIV rabbit polyclonal - and 10 fold excess of immunizing peptide for peptide competition experiments. After primary incubations, the cells were washed 3 times and incubated for 2 h at 4°C with 10 μ g/ml PE-labeled anti-rabbit antibody and/or 10 μ g/ml Alexa488-labeled EGF for simultaneous binding experiments. For Scatchard analysis the cells were incubated with various concentrations of Alexa488-EGF for a period of 7 hours (to ensure receptor trafficking steady state). Mean fluorescence levels were measured using a Coulter Epics XL flow cytometer (Beckman-Coulter) and further analyzed with WinMDI software (J. Trotter, Scripps University).

II.3 RESULTS

II.3.1 Rabbit polyclonal antisera generation

Two peptides were designed to mimic the EGFR domain II (DII) dimerization arm and one of domain IV (DIV) loops that is suspected to participate in receptor dimerization (Figure II.1). To best mimic their structure within EGFR, the peptides were synthesized as disulfide-bonded loops with additional spacer sequences (NH₂-KCGLYNPTTYQMDGCG-CONH₂ for DII and

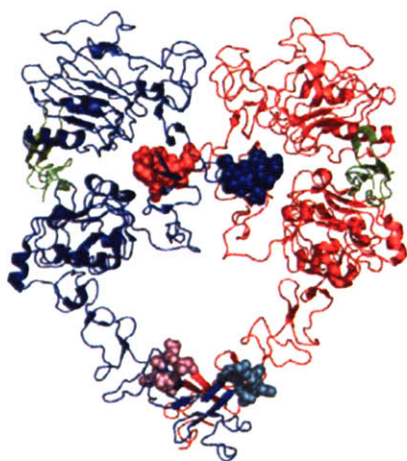


Figure II.1 – EGFR epitopes selected for peptide mimics and antibody harvesting. EGFR-EGF homodimer is shown as blue and red ribbons (EGFR) and green ribbons (EGFR). Epitopes from domain II dimerization arm (blue and red space-fill) and a putative dimerization loop in domain IV (cyan and magenta space fill) were selected for peptidomimetics and polyclonal antibody rising.

NH₂-CMGENNTC-CONH₂ for DIV). The peptides were then used to immunize rabbits, and polyclonal antisera against these peptides were harvested. Based on the locations of the immunizing peptides in the EGFR structure, it was expected that antibodies binding to these EGFR epitopes should sterically inhibit receptor dimerization.

Competition binding experiments in the presence of pAbs cognate peptide were performed to determine the specificity of each rabbit polyclonal antibody (pAb). Both anti-DII and anti-DIV pAbs showed strong

binding to the high EGFR expressing HMECs (Figure II.2). While cognate peptide competition resulted in reduced binding for both pAbs, the unrelated peptide did not have any effect in Ab binding. This demonstrates the specificity of the pAbs for the EGFR peptide epitopes they were raised to act on.

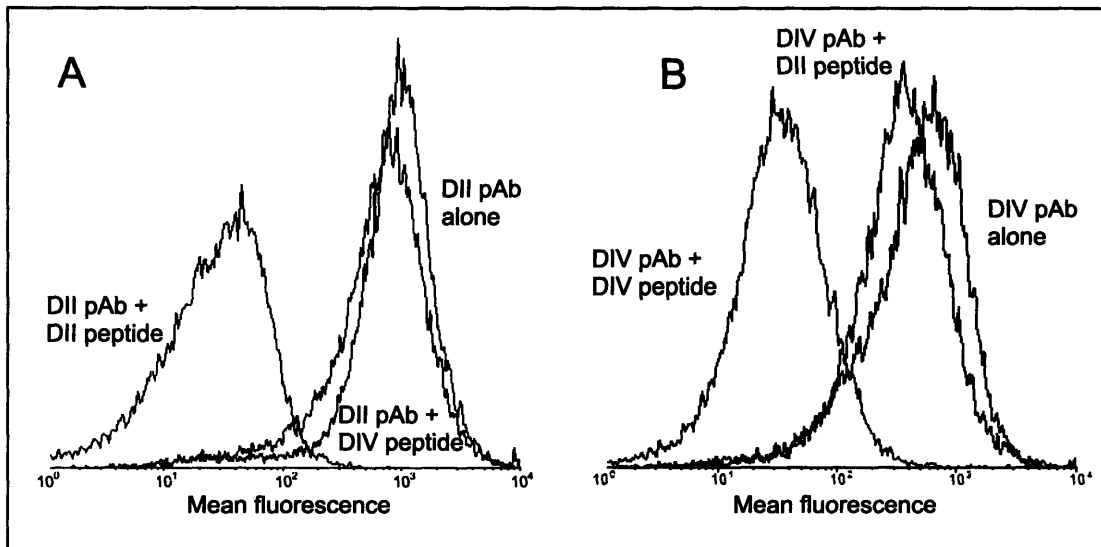


Figure II.2 – Specificity of rabbit polyclonal antibodies. The specificity of anti-DII (A) and anti-DIV (B) antibodies was tested by competition binding experiments. Antibody binding to EGFR on the surface of HMECs was tested alone and in the presence of cognate and cross peptides. Anti-DII antibody binding to EGFR was ablated by DII peptide, but not DIV peptide (A). Anti-DIV antibody binding to EGFR was ablated by DIV peptide, but not DII peptide (B).

II.3.2 Effect of antisera on receptor activation

In order to determine the effect of the pAbs on EGFR phosphorylation in HMECs we performed a series of quantitative western blots (Figure II.3). Upon treatment with either antibody the cells showed a concentration-dependent decrease in EGFR tyrosine 1173 phosphorylation. This response was specific to

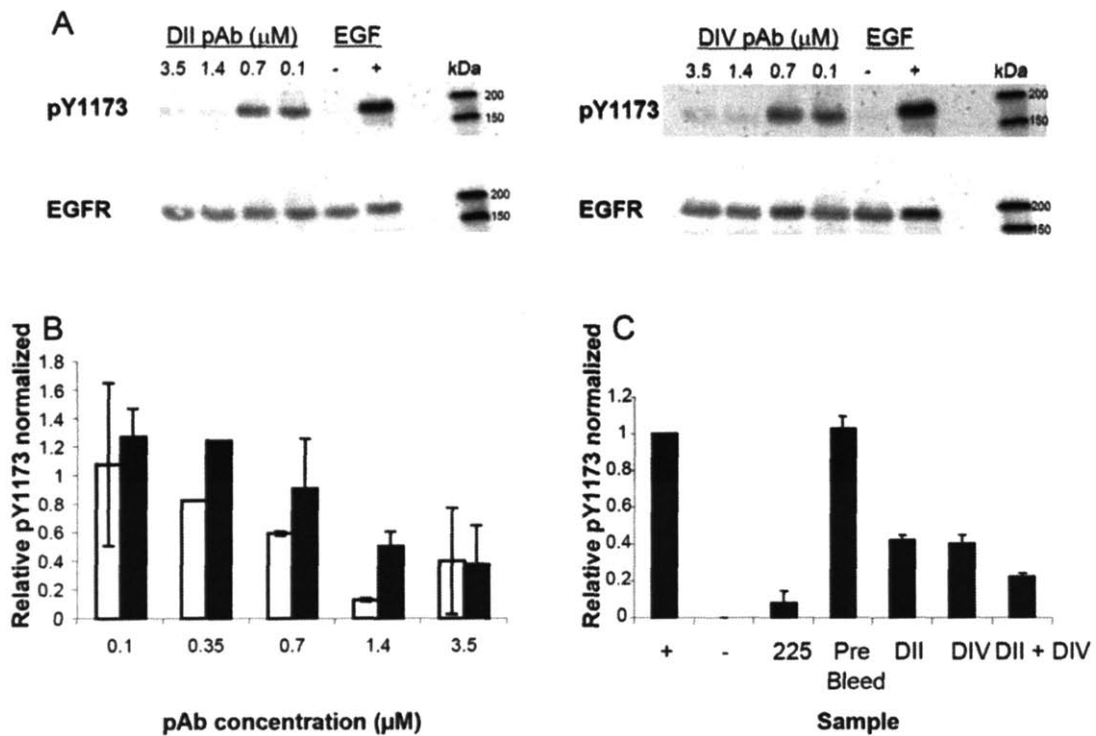


Figure II.3 – Effect of anti-DII and anti-DIV polyclonal antibodies in EGFR activity measured as pY 1173 auto-phosphorylation. (A) Exemplary quantitative western blot showing concentration dependent effect of anti-DII and anti-DIV on EGFR pY1173 auto-phosphorylation. **(B)** EGF binding mediated phosphorylation of EGFR pY1173 in the presence of anti-DII (white) and anti-DIV (black) antibodies normalized to EGF treatment only. At least 3 biological replicates were used, experiments as in (A). **(C)** Comparison of the effect on EGFR pY1173 phosphorylation of treating the cells with 100nM of monoclonal antibody 225 or 1.4μM of either pre-bleed serum, anti-DII, anti-DIV or an equal combination of anti-DII and anti-DIV.

antibodies binding to the peptides, as similar treatment with pre-injection antisera yielded no effect on EGFR phosphorylation (Figure II.3). The Abs also appear to act synergistically as a combination treatment of the cells with anti-DII and anti-DIV at the same total concentration as the Abs alone yielded a further inhibition of EGFR pY1173 phosphorylation than either Ab alone. However, ligand-competitive monoclonal antibody 225 performed better than both polyclonal, either alone or in combination (Figure II.3).

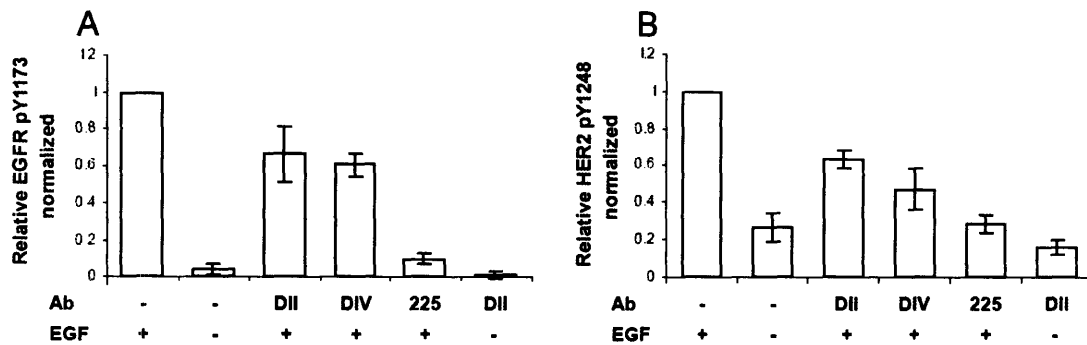


Figure II.4 – The polyclonal antibodies block homo- and hetero-dimerization. Cells were pretreated with 100nM of monoclonal antibody 225 or 1.4 μ M of anti-DII or anti-DIV antibodies.

(A) EGFR pY1173 relative phosphorylation was inhibited by both polyclonal antibodies and 225
(B) HER2 pY1248 relative phosphorylation was inhibited by both polyclonal antibodies and 225.
 These results indicate the ability of both polyclonal antibodies to block not only EGFR, but also EGFR mediated HER2 activation. Experiments were performed at least in triplicate.

Physiological levels of HER receptors in HMECs are 2×10^5 EGFR and 2×10^4 HER2 receptors per cell whereas HMECs 24H overexpress HER2 30 fold (6×10^5 receptors per cell) were used for further experiments to approximate a malignant state (Hendriks et al., 2003a). In order to assess the ability of our pAbs to block not only homo-, but also hetero-dimerization additional quantitative western blots were performed using pAb treatment on HMECs 24H (Figure II.4). The results show that both DII and DIV pAbs were able to reduce EGFR tyrosine 1173 and HER2 tyrosine 1248 phosphorylation in cells overexpressing HER2, implying that the pAbs inhibit both EGFR/EGFR homodimerization and EGFR/HER2 heterodimerization. Although when compared to monoclonal antibody (mAb) 225 (Sato et al., 1983) our antibodies show to be less effective in reducing receptor phosphorylation a direct comparison between the antibodies is not possible given the undefined affinity and composition of the polyclonal antisera. These experiments also demonstrated that the pAbs do not activate the

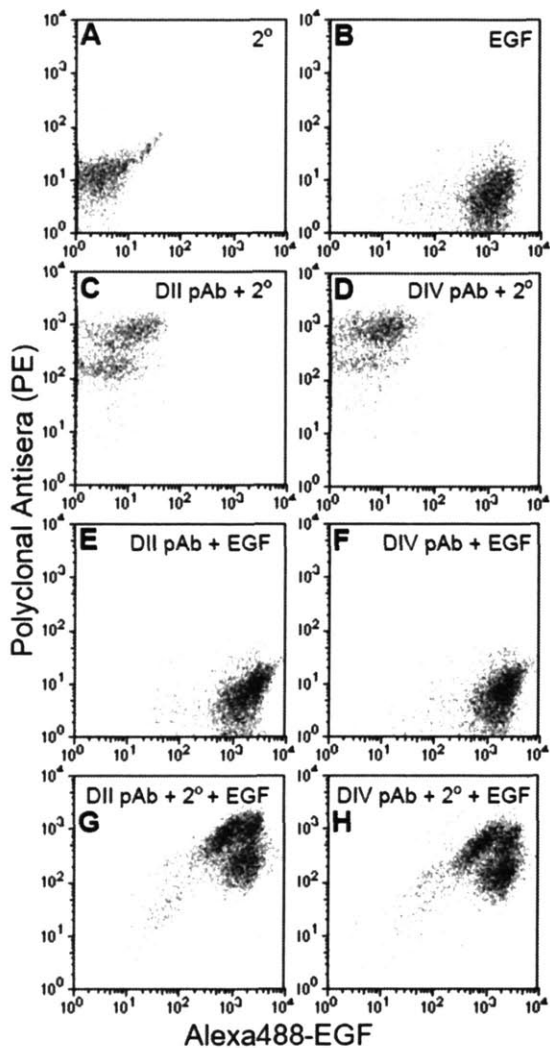


Figure II.5 – Binding of anti-DII and anti-DIV antibodies does not affect EGF binding to EGFR on HMECs. (A) Anti-rabbit Ab alone does not bind EGFR. **(B)** EGF binds EGFR. **(C)** Anti-DII binds EGFR. **(D)** Anti-DIV binds EGFR. **(E)** and **(G)** EGF binds EGFR in the presence of anti-DII. **(F)** and **(H)** EGF binds EGFR in the presence of anti-DIV.

receptor in the absence of EGF, a possibility suggested by preliminary data that was proven to be just an artifact (data not shown).

In order to discriminate the mechanism by which the pAb inhibited signaling, HMECs were labeled with the pAbs and a saturating concentration of Alexa488-labeled EGF. The cells were then analyzed by flow cytometry (Figure II.5). As shown in Figure II.5 the pAbs and EGF were able to bind cells simultaneously. Furthermore,

binding of pAb to the cells did not diminish the EGF binding to them. This demonstrates that the pAbs do not block EGF ligand binding and eliminates the possibility that

the pAbs were inhibiting receptor signaling by simply competing with ligand binding.

II.3.3 Effect of antisera on EGF binding

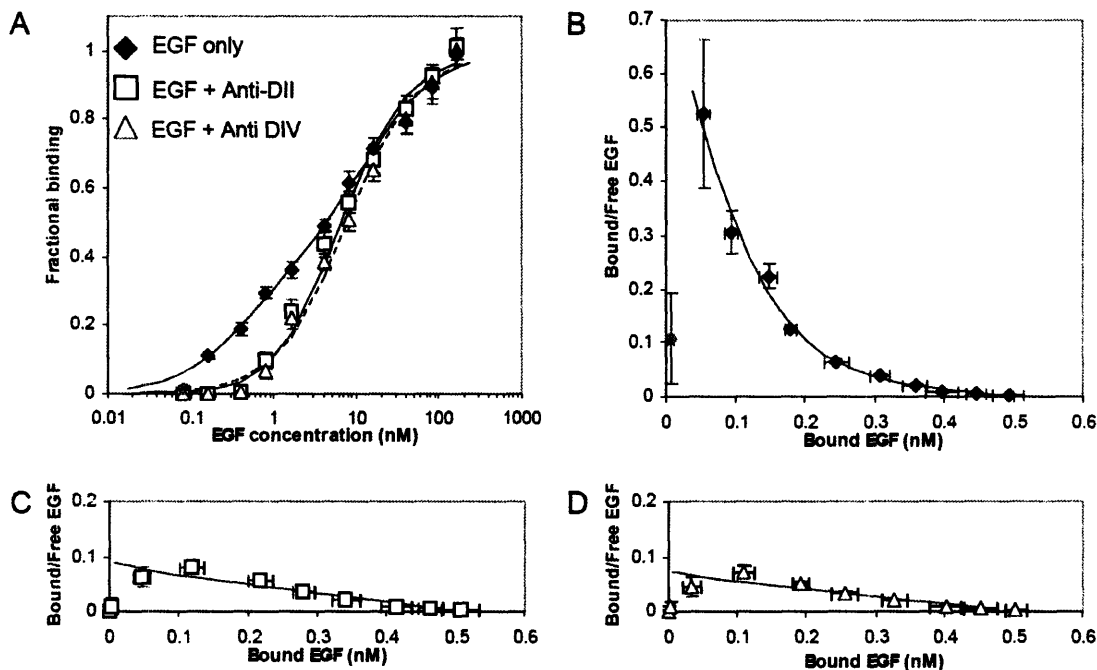


Figure II.6 – Equilibrium titration and Scatchard plots show that anti-DII and anti-DIV antibodies abrogate the high affinity binding component of EGF-EGFR. (A) Equilibrium titration of EGF-EGFR binding on the surface of HMECs. **◆** EGF only. **□** EGF in the presence of anti-DII. **△** EGF in the presence of anti-DIV. **(B), (C)** and **(D)** Corresponding Scatchard plots.

Although at saturating concentrations of EGF the pAbs did not affect EGF binding, pretreatment of HMECs with the pAbs resulted in altered EGF equilibrium titration and Scatchard plots (Figure II.6). In fact, while untreated cells titrated with EGF displayed the typical high affinity EGFR population seen previously in other cell lines (Mattoon et al., 2004; Schlessinger, 1988; Walker et al., 2004), which results in a concave-up Scatchard plot (Figure II.6B), pretreatment of the cells with either of the pAb eliminated this high-affinity

component, and the EGF titration data resembled a typical sigmoidal curve seen in single-site binding (Figure II.6A). Given that the elimination of high affinity EGF receptors resulted from the binding of either dimerization-blocking antibody we postulated that the high affinity component is due to the existence of preformed EGFR dimers. This hypothesis was further tested and computationally corroborated by Ginger Chao (Wolf Yadlin et al. work in progress)

II.4 DISCUSSION

In this chapter I have shown the effect of pAbs against EGFR domains II and IV in ligand mediated receptor phosphorylation, dimerization as well as ligand binding. Both pAbs (directed against domain II and IV) were shown to be specific for their target peptides and are able to block EGFR homodimerization and EGFR/HER2 heterodimerization, resulting in a significant decrease of EGFR and HER2 phosphorylation. Furthermore, inhibition of ligand binding can be ruled out as a reason for the observed reduced levels of receptor phosphorylation as both pAbs show no effect on ligand binding at the EGF saturating concentrations that were used for the phosphorylation assays. Whereas most anti-EGFR antibodies inhibit EGFR phosphorylation by competing with ligand binding (Graeven et al., 2006; Heymach et al., 2006; Holbro & Hynes, 2004; Sato et al., 1983; Yang et al., 2001), the anti-DII and anti-DIV antibodies inhibit EGFR mediated phosphorylation in a similar way to the way monoclonal antibody 2C4 (pertuzumab) inhibits HER2 mediated phosphorylation, i.e., by sterically blocking dimer formation (Franklin et al., 2004; Heymach et al., 2006; Holbro & Hynes, 2004).

To the extent of our knowledge this is the first reported case of anti-EGFR Abs acting through the inhibition of dimerization and independent of ligand competition. Another Ab, monoclonal antibody 806 has been shown to bind to a transitional form of EGFR between the tethered and extended monomeric

conformations (Johns et al., 2004; Walker et al., 2004), and its binding epitope and predicted structure suggest that binding of 806 should inhibit EGFR dimerization (Johns et al., 2004; Sivasubramanian et al., 2006). Whereas 806 acts (or doesn't) via inhibition of EGFR dimerization experimental results have shown that EGF binding can greatly decrease 806 binding to EGFR and hence its ability to inhibit EGFR activation in a physiological context (Walker et al., 2004). Therefore notwithstanding its mechanism of action 806 competes directly – by sterically impeding ligand receptor binding – or indirectly – by stabilizing an alternative receptor conformation that does not favor ligand binding – with EGF binding.

In the course of this work we have identified two novel EGFR domain II and IV epitopes. We raised rabbit polyclonal antibodies against these epitopes and showed that they inhibit receptor dimerization and activation while simultaneously allowing ligand binding. Although the pAbs do not inhibit EGF binding at saturating concentrations of the ligand, Figure II.6 shows that both pAbs eliminate the high affinity component of EGFR-EGF binding. This fact lead us to propose that this component is due to the presence of preformed EGFR dimers on the cell surface, a hypothesis that was computationally corroborated by a model developed by Ginger Chao. Furthermore, the model was able to satisfactorily fit the data without any assumption external to our current biological knowledge of EGFR biology (Wolf Yadlin et al).

Further refinement of antibodies with the same set of properties as above– for example humanized monoclonal antibodies against the domain II and domain

IV epitopes – may have very useful therapeutical applications as blocking agents of EFGR activated signaling cascades, specially in the treatment of tumors driven by autocrine EGFR signaling (Sizeland & Burgess, 1992). In this class of tumors ligand-competitive antibodies efficiency is low due to the high levels of effective ligand concentration on the surface of the cell, which make antibody blocking of ligand binding nearly impossible. A second potential target for the anti-DIV antibody may be EGFRvIII, a truncation mutant, which has lost expression of most of domain II – including the dimerization arm – and signals independently of ligand binding (Nishikawa et al., 1994). An anti-DIV Ab might inhibit EGFRvIII dimerization with itself or with wild type EGFR. Finally, antibodies binding to these domains II and IV epitopes could also be used in combination with each other or with other existing EGFR antibody and tyrosine kinase inhibitor therapies.

II.5 REFERENCES

Band V, Sager R (1989) Distinctive traits of normal and tumor-derived human mammary epithelial cells expressed in a medium that supports long-term growth of both cell types. *Proc Natl Acad Sci U S A* **86**: 1249-1253.

Franklin MC, Carey KD, Vajdos FF, Leahy DJ, de Vos AM, Sliwkowski MX (2004) Insights into ErbB signaling from the structure of the ErbB2-pertuzumab complex. *Cancer Cell* **5**: 317-328.

Graeven U, Kremer B, Sudhoff T, Killing B, Rojo F, Weber D, Tillner J, Unal C, Schmiegel W (2006) Phase I study of the humanised anti-EGFR monoclonal antibody matuzumab (EMD 72000) combined with gemcitabine in advanced pancreatic cancer. *Br J Cancer* **94**: 1293-1299.

Hendriks BS, Opresko LK, Wiley HS, Lauffenburger D (2003) Coregulation of epidermal growth factor receptor/human epidermal growth factor receptor 2 (HER2) levels and locations: quantitative analysis of HER2 overexpression

Holbro T, Hynes NE (2004) ErbB receptors: directing key signaling networks throughout life. *Annu Rev Pharmacol Toxicol* **44**: 195-217.

Janes KA, Albeck JG, Peng LX, Sorger PK, Lauffenburger DA, Yaffe MB (2003) A High-throughput Quantitative Multiplex Kinase Assay for Monitoring Information Flow in Signaling Networks: Application to Sepsis-Apoptosis. *Mol Cell Proteomics* **2**: 463-473.

Johns TG, Adams TE, Cochran JR, Hall NE, Hoyne PA, Olsen MJ, Kim YS, Rothacker J, Nice EC, Walker F, Ritter G, Jungbluth AA, Old LJ, Ward CW, Burgess AW, Wittrup KD, Scott AM (2004) Identification of the epitope for the epidermal growth factor receptor-specific monoclonal antibody 806 reveals that it preferentially recognizes an untethered form of the receptor. *J Biol Chem*: Epub 2004 Apr 2009.

Mattoon D, Klein P, Lemmon MA, Lax I, Schlessinger J (2004) The tethered configuration of the EGF receptor extracellular domain exerts only a limited control of receptor function. *Proc Natl Acad Sci U S A*: Epub 2004 Jan 2009.

Nishikawa R, Ji XD, Harmon RC, Lazar CS, Gill GN, Cavenee WK, Huang HJ (1994) A mutant epidermal growth factor receptor common in human glioma confers enhanced tumorigenicity. *Proc Natl Acad Sci U S A* **91**: 7727-7731.

Sato JD, Kawamoto T, Le AD, Mendelsohn J, Polikoff J, Sato GH (1983) Biological effects in vitro of monoclonal antibodies to human epidermal growth factor receptors. *Mol Biol Med* **1**: 511-529.

Satoh T, Kato J, Nishida K, Kaziro Y (1996) Tyrosine phosphorylation of ACK in response to temperature shift-down, hyperosmotic shock, and epidermal growth factor stimulation. *FEBS Lett* **386**: 230-234.

Schlessinger J (1988) The epidermal growth factor receptor as a multifunctional allosteric protein. *Biochemistry* **27**: 3119-3123.

Sivasubramanian A, Chao G, Pressler HM, Wittrup KD, Gray JJ (2006) Structural model of the mAb 806-EGFR complex using computational docking followed by computational and experimental mutagenesis. *Structure* **14**: 401-414.

Sizeland AM, Burgess AW (1992) Anti-sense transforming growth factor alpha oligonucleotides inhibit autocrine stimulated proliferation of a colon carcinoma cell line. *Mol Biol Cell* **3**: 1235-1243.

Stampfer MR, Bartley JC (1985) Induction of transformation and continuous cell lines from normal human mammary epithelial cells after exposure to benzo[a]pyrene. *Proc Natl Acad Sci U S A* **82**: 2394-2398.

Walker F, Orchard SG, Jorissen RN, Hall NE, Zhang HH, Hoyne PA, Adams TE, Johns TG, Ward C, Garrett TP, Zhu HJ, Nerrie M, Scott AM, Nice EC, Burgess AW (2004) CR1/CR2 interactions modulate the functions of the cell surface epidermal growth factor receptor. *J Biol Chem*: Epub 2004 Mar 2011.

Yang XD, Jia XC, Corvalan JR, Wang P, Davis CG (2001) Development of ABX-EGF, a fully human anti-EGF receptor monoclonal antibody, for cancer therapy. *Crit Rev Oncol Hematol* **38**: 17-23.

**III TIME-RESOLVED MASS SPECTROMETRY OF
TYROSINE PHOSPHORYLATION SITES IN THE EGF
RECEPTOR SIGNALING NETWORK REVEALS
DYNAMIC MODULES**

III.1 SUMMARY

Ligand binding to cell surface receptors initiates a cascade of signaling events regulated by dynamic phosphorylation events on a multitude of pathway proteins. Quantitative features, including intensity, timing, and duration of phosphorylation of particular residues, may play a role in determining cellular response, but experimental data required for analysis of these features have not been available before this thesis work.

In order to understand the dynamic operation of signaling cascades, we have developed a methodology enabling the simultaneous quantification of tyrosine phosphorylation of specific residues on dozens of key proteins in a time-resolved manner, downstream of EGFR activation. Although we used EGFR signaling as a proof of principle system, our methodology can be easily extended for its use in alternative signaling cascades and to compare dynamic and static phosphorylation states of different biological samples including clinical biopsies.

Cells were stimulated with EGF for 0, 5, 10 and 30 minutes and tryptic peptides from each stimulation time point were labeled each with one of four isoforms of the iTRAQ reagent to enable downstream quantification. The labeled samples were mixed and tyrosine-phosphorylated peptides were immunoprecipitated with an anti-phosphotyrosine antibody and further enriched by IMAC prior to LC/MS/MS analysis. We identified and quantified relative time courses of 78 tyrosine phosphorylation sites on 58 proteins from a single analysis. Replicate analyses of a separate biological sample provided both

validation of this first data set and identification of 26 additional tyrosine phosphorylation sites and 18 additional proteins.

Computational analysis of the quantitative temporal phosphorylation profiles using self-organizing maps resulted in identification of several groups of tyrosine residues exhibiting similar temporal phosphorylation profiles. Each group defined a dynamic module in the EGFR signaling network consistent in many cases with particular cellular processes and/or potential phosphorylation by a limited number of related tyrosine kinases. The presence of novel proteins and associated tyrosine phosphorylation sites within these modules indicates additional components of the EGFR network and potentially localizes the topological action of these proteins.

Part of this work has been previously published in molecular cellular proteomics (MCP) volume 4(9), September 2005, pages 1240-1250 (Zhang et al., 2005c). MCP does not require copyright permission request for students wanting to reproduce or republish their work for educational purposes, please refer to the permission section at the end of this thesis or in its defect please refer to http://www.mcponline.org/misc/Copyright_Permission.shtml.

III.2 MATERIALS AND METHODS

III.2.1 Cell Culture, EGF stimulation

184A1 HMECs (parental human mammary epithelial cells (Stampfer & Bartley, 1985)) were a kind gift from Martha Stampfer (Lawrence Berkeley Laboratory, Berkeley, CA) and were maintained in DFCI-1 medium supplemented with 12.5 ng/ml EGF, as in (Hendriks et al., 2003a). Cells were washed with PBS, and incubated for 12 hours in serum free media (DFCI-1 without EGF, bovine pituitary extract, or fetal bovine serum) after 80% confluence was reached in 15 cm plates (~ 3×10^7 cells). The synchronized cells were washed with PBS after removal of media. The cells were then stimulated with 25 nM EGF in serum free media without sodium bicarbonate for 5, 10 or 30 minutes, or left untreated with serum free media for 5 min as control before lysing.

III.2.2 Protein Quantification

Protein concentration of cell lysates was determined using the micro bicinchoninic acid assay (micro-BSA – Pierce) according to the manufacturer's protocol.

III.2.3 Cell Lysis, Protein Digestion and Peptide Fractionation

After EGF stimulation, cells were lysed on ice with 3 ml of 8 M urea supplemented with 1 mM Na₃VO₄. A 10 µl aliquot was taken from each sample to perform the micro bicinchoninic acid protein concentration assay (Pierce) according to the manufacturer's protocol. Cell lysates were reduced with 10 mM DTT for 1hr at 56 °C, alkylated with 55 mM iodoacetamide for 45 min at room temperature, and diluted to 12 ml with 100 mM ammonium acetate, pH 8.9. 40 µg trypsin (Promega) was added to each sample (~ 100:1 substrate to trypsin ratio) and the lysates were digested overnight at room temperature. The whole cell digest solutions were acidified to pH 3 with acetic acid (HOAc), and loaded onto C18 Sep-Pak Plus Cartridges (Waters). The peptides were desalted (10 mL 0.1% HOAc) and eluted with 10 mL of a solution comprised of 25% acetonitrile (MeCN) with 0.1% HOAc. Each sample was divided into ten aliquots and lyophilized overnight to dryness for storage at -80°C.

III.2.4 iTRAQ Labeling and peptide Immunoprecipitation

Peptide labeling with iTRAQ reagent (Applied Biosystems) was performed according to the manufacturer's protocol. Briefly, each aliquot (3 ×10⁶ cell equivalent) was reacted with one tube of iTRAQ reagent (e.g. acquisition of this data set from the first biological sample required 8 tubes of iTRAQ reagent – 2 x

iTRAQ-114 (0 min), 2 x iTRAQ-115 (5 min), 2 x iTRAQ-116 (10 min), and 2 x iTRAQ-117 (30 min)) after the sample was dissolved in 30 μ l of 0.5 M triethylammonium bicarbonate ($\text{N}(\text{Et})_3\text{HCO}_3$) pH 8.5 and the iTRAQ reagent was dissolved in 70 μ l of ethanol. The mixture was incubated at room temperature for 1 hr and concentrated to \sim 20 μ l. Samples labeled with four different isotopic iTRAQ reagents were combined and concentrated to 10 μ l, then dissolved in 200 μ l of IP buffer (100 mM Tris, 100 mM NaCl, 1% NP-40, pH 7.4) and 200 μ l of water, and pH was adjusted to 7.4. The mixed sample was incubated with 4 μ g immobilized PY99 (Santa Cruz) overnight at 4OC. The antibody beads were spun down for 5 minutes at 7000 rpm and the supernatant was separated and saved. The antibody-bound beads were washed twice with 200 μ l IP buffer for 10 minutes and twice with rinse buffer (100 mM Tris, 100 mM NaCl, pH 7.4) for 5 minutes at room temperature. The phosphotyrosine containing peptides were eluted from antibody with 50 μ l of 100 mM glycine pH 2.5 for 1 hr at room temperature.

III.2.5 IMAC and Mass Spectrometry Analysis

Phosphopeptide enrichment on IMAC was performed as described (Ficarro et al., 2002), with the exception that peptides were not converted to methyl esters. Following antibody elution, peptides were loaded onto a 10-cm long self-packed IMAC (20MC, Applied Biosystems) capillary

column (200 μm ID, 360 μm OD), and rinsed with organic rinse solution (25% MeCN, 1% HOAc, 100 mM NaCl) for 10 min at 10 $\mu\text{L}/\text{min}$. The column was equilibrated with 0.1% HOAc for 10 min at 10 $\mu\text{L}/\text{min}$ and eluted onto a 10-cm long self-packed C18 capillary precolumn (100 μm ID, 360 μm OD) with 50 μl of 250 mM Na_2HPO_4 , pH 8.0. After 10 min rinse (0.1% HOAc), the precolumn was connected to a 10-cm long self-packed C18 (YMC-Waters 5 μm ODS-AQ) analytical capillary column (50 μm ID, 360 μm OD) with an integrated electrospray tip (~ 1 μm orifice). Peptides were eluted using a 100 min gradient with solvent A ($\text{H}_2\text{O}/\text{HOAc}$, 99/1 vol/vol) and B ($\text{H}_2\text{O}/\text{MeCN}/\text{HOAc}$, 29/70/1 vol/vol): 10 min from 0% to 15% B, 75 min from 15% to 40% B, 15 min from 40% to 70% B. Eluted peptides were directly electrosprayed into a quadrupole time-of-flight mass spectrometer (QSTAR XL Pro, Applied Biosystems). MS/MS spectra of five most intense peaks with 2 to 5 charges in the MS scan were automatically acquired in information-dependent acquisition with previously selected peaks excluded for 40 seconds.

III.2.6 Western Blot Analysis

40 μg protein from each sample was mixed with 4x sample buffer (250 mM Tris-HCl pH 6.8, 8% SDS, 40% glycerol, 0.04% bromophenol blue and 400 mM DTT) and boiled for 5 minutes. Proteins were separated by SDS-PAGE and transferred onto PDVF membranes. After blocking for 1 hour at room temperature

membranes were incubated overnight at 4°C in primary antibody, washed 3 times for 5 min in TBS-T (20 mM Tris-HCl pH 7.5, 137 mM NaCl, 0.1% Tween-20), incubated for 1 hr at room temperature in secondary antibody (dilution 1:2500 Horseradish peroxidase conjugated donkey anti-rabbit in TBS-T, 5% non-fat milk powder) (Amersham Biosciences) and finally washed 3 times for 5 min with TBS-T. Blots were developed with ECL Advance western blotting detection kit (Amersham) and scanned on a Kodak Image Station 1000. Primary antibodies used were anti-EGFR, for loading control; anti-EGFR pY1148; anti-FAK pY576; anti-STAT-3 pY705 (all from Cell Signaling Technologies) and anti-Gsk-3- β pY216 (Upstate Biotechnology).

III.2.7 Phosphopeptide Sequencing, Data Clustering and Analysis

MS/MS spectra were extracted and searched against human protein database (NCBI) using ProQuant (Applied Biosystems) and MASCOT (Matrix Science). For ProQuant, an interrogator database was generated by predigesting the human protein database with trypsin and allowing one miscleavage and up to six modifications on a single peptide (phosphotyrosine \leq 2, phosphoserine \leq 1, phosphothreonine \leq 1, iTRAQ-lysine \leq 4, and iTRAQ-tyrosine \leq 4). Mass tolerance was set to 0.15 amu for precursor ions and 0.1 amu for fragment ions. For MASCOT, data was searched against the human non-redundant protein database with trypsin specificity, 2 missed cleavages, precursor mass tolerance

set to 1.5 Da and fragment ion tolerance set to 0.2 Da. Phosphotyrosine-containing peptides were manually validated and quantified. Peak areas for each of the four signature peaks (m/z: 114, 115, 116, 117) were obtained and corrected according to the manufacturer's instructions to account for isotopic overlap. Data was further corrected with values generated from the peak areas of non-phosphorylated peptides to account for possible variations in the starting amounts of sample for each time point. Finally, all the data was normalized by the 5 minute sample. Mean phosphorylation, standard deviation and p- values to estimate statistical significance for differential phosphorylation between the different time points were calculated using Excel. The p-values were calculated using a paired, two-tailed student test. A self organizing map was generated with the Spotfire program to cluster phosphorylation sites with self-similar profiles. All the analyses using Spotfire were done with original built in functions of the program.

III.3 RESULTS

III.3.1 Time Resolved Tyrosine Phosphorylation after EGF Stimulation Measured by Quantitative Mass Spectrometry

III.3.1.1 Methodology

Stable-isotope labeling either during cell culture or after cell lysis provides a simple and accurate method for quantifying proteins and peptides from different cell states (Ong et al., 2002; Ross et al., 2004b). Stable-isotope-tagged amine-reactive reagents (iTRAQ) from Applied Biosystems enable the comparison of up to four cell states in a single analysis (Ross et al., 2004b). By combining iTRAQ-based labeling of different cellular samples with immunoprecipitation of tyrosine-phosphorylated peptides (Rush et al., 2004) we have generated a new method enabling quantitative analysis of specific tyrosine phosphorylation sites in four populations of cells. In this study we subjected the cells to EGF stimulation for 0, 5, 10 and 30 minutes. For each time point, a single 15 cm plate of cells was cultured under the same conditions prior to stimulation and EGF (25 nM) was added for the appropriate length of time. After cell lysis the proteins from each sample were digested with trypsin and desalted. Aliquots corresponding to 20% (~1.5x10⁶ cell equivalents / ~400µg of peptides) of each one of the resulting four peptides set were labeled with iTRAQ reagent, mixed together, and immunoprecipitated with anti-phosphotyrosine antibody (PY99 – Santa Cruz

Biotechnologies). Following immunoprecipitation the peptides were further enriched by immobilized metal affinity chromatography (IMAC) prior to LC-MS/MS analysis (Figure III.1A). IMAC enrichment was a necessary to clean the sample from non-specific binding of peptides to both the antibody and/or the protein G conjugated agarose beads. Because iTRAQ consist of four isobaric markers – indistinguishable from each other in MS mode – peptides tagged with the different forms of the reagent co-elute during the LC gradient and generate a single peak for each charge state in the MS scan (Ross et al., 2004b). Following selection of a peak in MS mode, MS/MS fragmentation of the iTRAQ labeled peptide results in 4 signatures peaks at m/z 114, 115, 116 and 117 respectively (Figure III.1B and III1C), while fragmentation along the peptide backbone results in b- and y-type fragment ions which may be used to identify the peptide sequence (Figure III.1B). In order to minimize noise-associated quantification errors and facilitate comparison of the temporal phosphorylation profiles, each of the phosphorylation profiles was normalized relative to the 5 minute time point, as this time point typically provided the greatest signal-to-noise ratio in our system. To normalize the results taking into account protein level and labeling efficiency, the supernatant from the anti-pY peptide immunoprecipitation was analyzed and iTRAQ marker ions from non-phosphorylated peptides were averaged and used to correct the data (Figure III.1C).

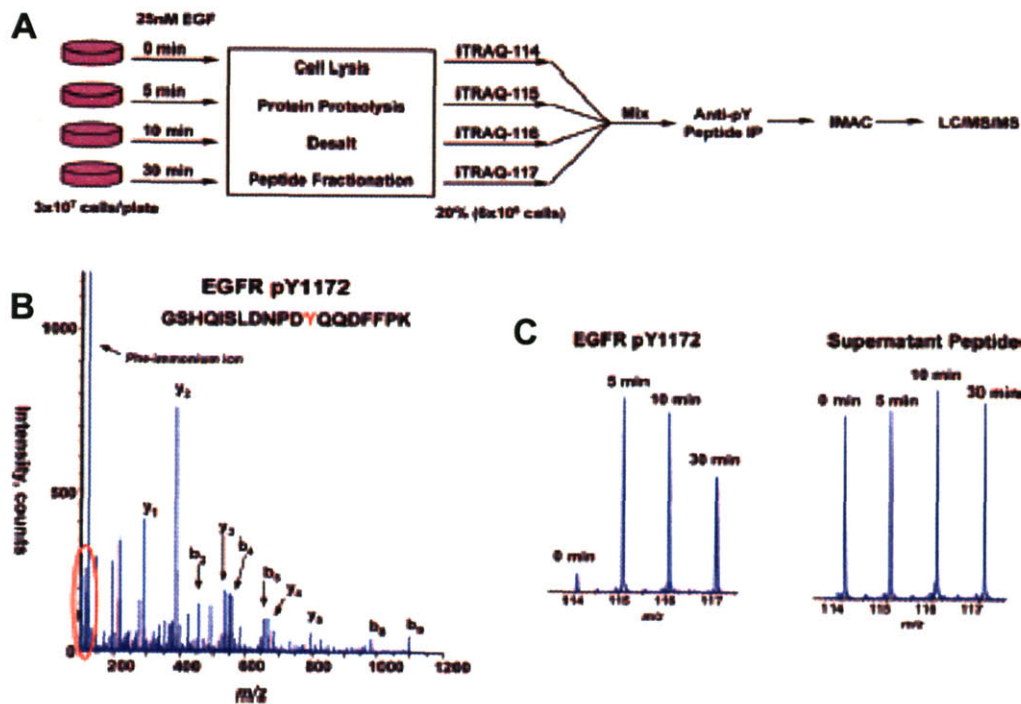


Figure III.1 – Mass spectrometry-based approach to quantitative analysis of phosphorylation dynamics on specific tyrosine residues. (A) Four plates of cells were cultured under normal conditions, serum starved for 12 hours, and stimulated with 25 nM EGF for 0, 5, 10, or 30 minutes. Following cell lysis, proteins were proteolyzed and the resulting peptide mixture desalted prior to aliquoting into ten equivalent fractions per stimulation time. For each stimulation time, one aliquot was labeled with one of the four isoforms of the iTRAQ reagent and then mixed with the labeled aliquots from the other time points. Phosphotyrosine containing peptides were immunoprecipitated and further enriched by IMAC affinity chromatography prior to LC/MS/MS analysis. (B) For each MS/MS spectrum, γ - and b -type fragment ions (containing the C- and N-terminus of the peptide, respectively) enable the identification of the peptide sequence including the specific localization of the phosphorylated residue. (C) Quantitative information is encoded in the low mass-to-charge ratio portion of the MS/MS spectrum (highlighted in (B), red oval). The 0 minute time point was labeled with iTRAQ-114, the 5 minute time point was labeled with iTRAQ-115, the 10 minute time point was labeled with iTRAQ-116, and the 30 minute time point was labeled with iTRAQ-117. Relative peak areas of the 4 marker ions were used to quantify tyrosine phosphorylation levels. Protein levels and labeling efficiency were normalized with quantification data generated by analysis of peptides contained in the supernatant following peptide IP. The marker ion region of a representative non-phosphorylated peptide is on the right side of (C).

By applying this methodology to investigate EGFR signaling in parental HMEC cells we were able to identify 78 tyrosine phosphorylation sites on 58 proteins in a single analysis. For each phosphorylation site, a quantitative

Table III.1 Quantitative Time Course Tyrosine Phosphorylation Profiles.

A subset of protein phosphorylation sites identified from HMEC cells treated with EGF for varying duration. Results are presented from analysis of two different cell cultures, for the second cell culture analytical replicates were also performed; the average value is presented in the table. For each protein, specific phosphorylation site(s) were identified and quantified by comparing relative peak areas of the iTRAQ marker ions. The 5 minute sample was used to normalize the phosphorylation profiles. The data is therefore presented as a ratio relative to the 5 minute time point. Statistical significance was calculated using a paired, two-tailed student test.

ID		Quantification: Sample 1, analysis 1			Sample 2, average of analysis 1 and 2		
Protein	Site	Ratio 0 min:5 min	Ratio 10 min:5 min	Ratio 30 min:5 min	Ratio 0 min:5 min	Ratio 10 min:5 min	Ratio 30 min:5 min
c-Cbl	Tyr-552	0.000 ± 0.001***	0.918 ± 0.015	0.385 ± 0.005**	0.028 ± 0.008***	0.715 ± 0.016***	0.314 ± 0.016***
Caveolin 1	Tyr-14	0.391 ± 0.074	1.034 ± 0.009	0.792 ± 0.037	0.425 ± 0.027***	1.121 ± 0.136	0.804 ± 0.061***
EGFR	Tyr-1172	0.050 ± 0.031*	0.900 ± 0.019	0.505 ± 0.012*	0.039 ± 0.021***	0.945 ± 0.109	0.662 ± 0.035***
EGFR	Tyr-1092	0.059	0.744	0.429	0.034 ± 0.016***	0.689 ± 0.111*	0.461 ± 0.095**
EphB1	Tyr-600	0.656 ± 0.090	1.142 ± 0.045	1.064 ± 0.049	0.733 ± 0.038	1.070 ± 0.009	1.182 ± 0.003**
EphrinB2	Tyr-304	0.837	1.364	1.156	0.792 ± 0.153	0.902 ± 0.234	1.072 ± 0.190
EPS15	Tyr-849	0.080	1.409	0.652	0.043 ± 0.050*	1.412 ± 0.130	0.605 ± 0.046
FAK	Tyr-576	0.637 ± 0.051***	1.424 ± 0.051***	1.629 ± 0.036***	0.569 ± 0.088***	0.801 ± 0.076***	1.678 ± 0.206***
FRK	Tyr-497	0.814	1.241	1.340	0.495 ± 0.050*	1.349 ± 0.053	0.773 ± 0.037
GIT 1	Tyr-545	0.575 ± 0.057	1.183 ± 0.016*	1.093 ± 0.040	0.807 ± 0.083*	1.212 ± 0.276	1.074 ± 0.328
GSK-3-β	Tyr-279	0.846 ± 0.014**	1.112 ± 0.024*	0.931 ± 0.028	0.908 ± 0.115	1.147 ± 0.122	0.980 ± 0.082
HER2	Tyr-1248	0.141	0.732	0.532	0.214	0.808	0.492
p130Cas	Tyr-327	1.126	1.136	1.442	2.116 ± 0.130**	1.418 ± 0.056**	1.497 ± 0.045**
ACK1	Tyr-857	0.133	0.997	0.329	0.099 ± 0.037*	0.561 ± 0.069	0.356 ± 0.109
Paxillin	Tyr-116	0.528 ± 0.131	0.755 ± 0.137	0.792 ± 0.164	0.726 ± 0.045***	0.732 ± 0.105**	0.850 ± 0.095*
Plakophilin 3	Tyr-84	0.194	0.778	0.144	0.036 ± 0.007**	0.429 ± 0.015*	0.286 ± 0.023*
PLC-γ	Tyr-775, Tyr-783	0.043	0.605	0.335	0.008 ± 0.016**	0.287 ± 0.013**	0.231 ± 0.001***
PTP NR 11	Tyr-62	0.945 ± 0.032	1.266 ± 0.023*	1.122 ± 0.016	1.074 ± 0.157	0.971 ± 0.142	1.128 ± 0.202
PTPR A	Tyr-798	0.833 ± 0.005*	0.943 ± 0.010	0.708 ± 0.003**	0.873 ± 0.073*	0.740 ± 0.119*	0.613 ± 0.103**
Rho GEF 5	Tyr-19, Tyr-22	0.019 ± 0.037*	0.844 ± 0.052	0.832 ± 0.047	0.000 ± 0.012**	0.352 ± 0.012**	0.248 ± 0.040*
SHC	Tyr-238, 240	0.147 ± 0.015**	1.160 ± 0.083	0.902 ± 0.059	0.173 ± 0.023*	0.995 ± 0.174	0.734 ± 0.198
SHC	Tyr-318	0.028 ± 0.046*	1.077 ± 0.026	0.846 ± 0.003**	0.296 ± 0.392	0.878 ± 0.217	0.908 ± 0.543
SHIP-2	Tyr-886	0.043 ± 0.054***	0.801 ± 0.080	0.363 ± 0.028**	0.013 ± 0.045***	0.428 ± 0.160***	0.291 ± 0.160**
STAM-1	Tyr-381	0.309	2.659	0.926	0.146 ± 0.288*	4.188 ± 0.724**	1.624 ± 0.397
STAM-2	Tyr-374	0.218	2.753	0.908	0.459 ± 0.474	2.515 ± 0.613*	1.308 ± 0.017***
STAT-3 Isoform 1	Tyr-705	0.209 ± 0.026***	0.803 ± 0.073*	0.244 ± 0.046***	0.242 ± 0.092**	0.565 ± 0.039**	0.478 ± 0.134*
STAT-3 Isoform 2	Tyr-704	0.195	0.794	0.275	0.196 ± 0.197	0.545 ± 0.178	0.406 ± 0.266

*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

temporal phosphorylation profile was generated by comparing the relative ratios of peak areas for the iTRAQ marker ions (m/z 114-117) in the MS/MS spectrum.

Table 1 provides a subset of mean values, standard deviations and an estimation

of the statistical significance of the change in phosphorylation between each time point relative to the five minutes sample for several biologically interesting tyrosine phosphorylation sites. For example, the EGFR pY1148 autophosphorylation site increased 20-fold from 0 to 5 minutes, decreased by 10% from 5 to 10 minutes, and decreased by another 40% to 30 minutes. Because a typical feature of IDA is that many of the peptides are selected for MS/MS more than once during the chromatographic elution profile, resulting in multiple independent measurements of the relative iTRAQ marker ion peak areas ratios in the LC/MS/MS analysis, we can confidently affirm that the difference between the 0 minute and 5 minute time point and the 30 minute and 5 minute time point is statistically significant ($p < 0.05$), whereas the difference between the 10 minute and 5 minute time point does not meet this significance criterion for this particular phosphorylation site. This kind of information is important to distinguish between real changes and apparent changes between the samples.

III.3.1.2 Protein Phosphorylation Insights

Of the total 58 proteins we identified in this analysis, 52 have been previously associated with the EGFR signaling network, while the other six proteins (listed in Table 2) have not been previously identified in either proteomic or biochemical analyses of EGFR signaling. Representative spectra for two of the six peptides listed in Table III.2 are presented in Figure III.2; the mass range spanning the iTRAQ marker ion region is highlighted for each peptide. Included

Table III.2 Tyrosine-phosphorylated proteins not previously reported in EGFR signaling
 Of the 58 proteins identified in the analysis of the first biological sample, six have not been previously associated with the EGFR signaling network. Protein identification, tyrosine phosphorylation site, and time course profile data are provided for these six proteins.

Protein	Site	Ratio		
		0 min:5 min	10 min:5 min	30 min:5 min
Chromosome 20 ORF 18 (HOIL-1)	pY288	0.026 ± 0.034***	1.950 ± 0.156**	0.441 ± 0.012***
Hypothetical protein FLJ14564	pY413	0.66 ± 0.04	0.956 ± 0.063	0.719 ± 0.023
Hypothetical protein FLJ30532 (MARVELD2)	pY23	0.000	0.753	0.534
Retinoic acid induced 3	pY320	0.725 ± 0.117	1.177 ± 0.027	0.891 ± 0.029
KIAA0606 protein	pY2060	0.140	0.671	0.413
KIAA1217 protein	pY393	0.456	1.102	0.952

*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

in this group of six proteins is KIAA0606, a protein also known as suprachiasmatic nucleus circadian oscillatory protein (SCOP). Regulation of the circadian clock in the suprachiasmatic nucleus drives daily rhythms of behavior and has been associated with EGF receptor signaling in the hypothalamus (Kramer et al., 2003). Additionally, EGF activated expression of clock genes Per1 and Per2 and modulation of the circadian rhythm in cell culture has also been documented (Balsalobre et al., 2000). Although the mechanism by which signal is transduced from the EGF receptor to the clock gene promoter region has not been elucidated, the presence of SCOP in this data set might serve as a starting point to answer this question. In fact we could hypothesize that the detected rapid and transient pY2060 SCOP phosphorylation in response to EGF stimulation that we found in this work might be directly involved in signal transduction from activated EGF receptor to downstream upregulation of clock gene expression. In any case the hypothesis remains that and this relationship needs to be biologically validated. One of the strengths of our method is in

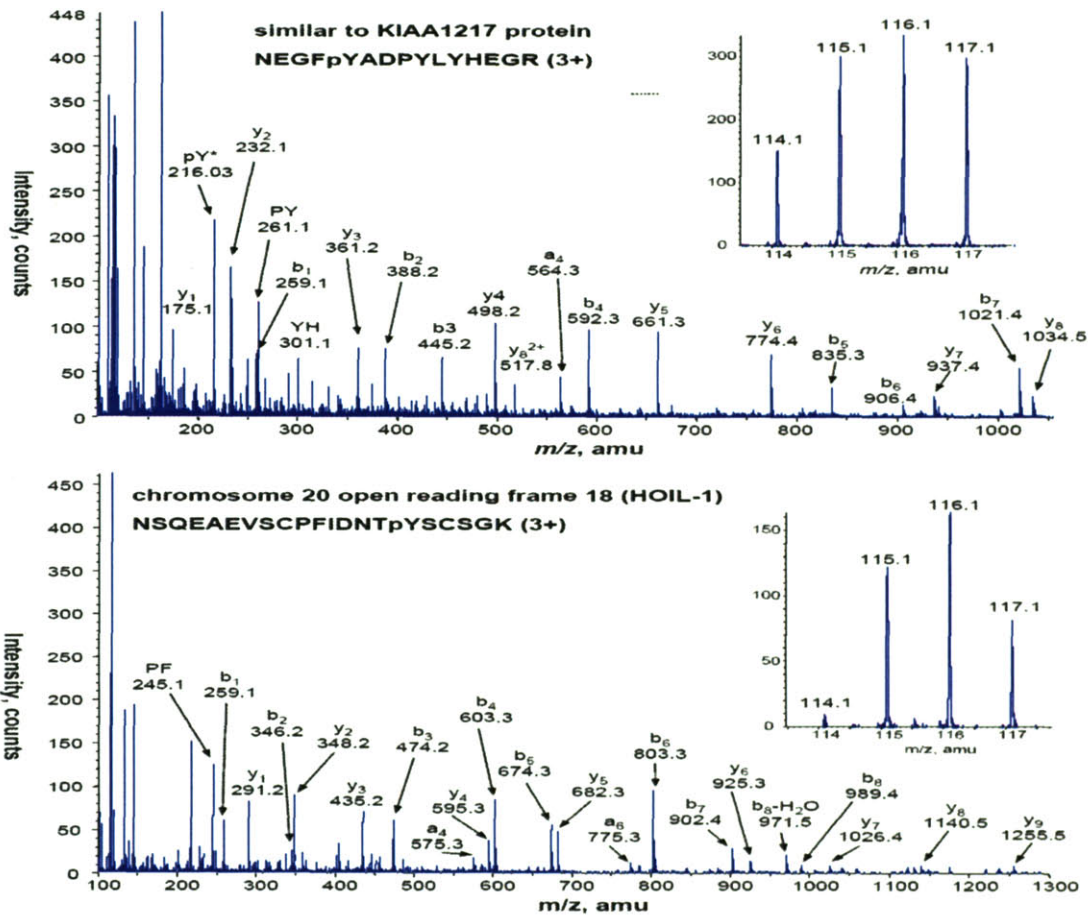


Figure III.2 – Representative MS/MS spectra for two phosphorylated peptides from proteins not previously reported in EGFR signaling. For each spectrum, y- and b-type fragment ions present in full scan mass spectra enable peptide identification and phosphorylation site assignment, while peak areas for each of the iTRAQ marker ions (inset for top and bottom) enable quantification of the temporal phosphorylation profiles.

highlighting this and many other hypotheses that other methodologies would overlook, but it does not replace experimental confirmation.

Another interesting example of the previously EGFR-unrelated protein group is KIAA1217, a protein whose function has not been characterized to date. Notably, the specific peptide and phosphorylation site we have identified on this protein display surprising homology (12/15 identical residues) to a region of p130Cas interacting protein (p140Cap), a protein known to be tyrosine

Table III.3 – Tyrosine Phosphorylation Sites not previously reported in the literature.

In addition to the sites listed in Table 2, many additional tyrosine phosphorylation sites which have not been previously reported in the literature have been identified in this work. A subset of these sites is presented along with the protein identification and the temporal phosphorylation profile quantification.

Protein	Site	Ratio		
		0 min:5 min	10 min:5 min	30 min:5 min
Hypothetical protein FLJ00269	pY411	0.808	0.841	1.026
Hypothetical protein FLJ21610	pY453	0.261	1.123	0.888
Chromosome 3 ORF 6 (Ymer)	pY145	0.00 ± 0.003***	1.359 ± 0.024*	0.444 ± 0.005**
Target of myb1-like 2	Tyr-160 & Tyr-168	0.059 ± 0.023*	0.640 ± 0.152	0.249 ± 0.058*

*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

phosphorylated in response to stimulation with exogenous growth factors and that probably plays a role in cell motility (Di Stefano et al., 2004).

Not only have we identified these 6 novel proteins as potential part of the EGFR network, we also identified several novel phosphorylation sites on proteins known to be in the network (a subset of these sites is presented in Table 3). In this group we can find phosphorylation sites on hypothetical protein FLJ00269, hypothetical protein FLJ21610, target of myb1-like 2 protein, and chromosome 3 open reading frame 6 (previously named Ymer by Blagoev et al) (Blagoev et al., 2004). Identification and quantification of temporal phosphorylation profiles at specific tyrosine sites in these proteins further confirms the involvement of these proteins in the EGFR signaling cascade. Furthermore, knowledge of the specific phosphorylation sites on these and other proteins that participate in EGFR mediated signaling should facilitate their biological manipulation, helping to characterize their functional role in the EGFR network – for example by doing Y→F mutation experiments in these particular sites and assessing their implication in downstream phosphorylation and cell phenotype .

III.3.1.3 Validation

To validate our methodology and our first data set in particular, a second set of cells were cultured, stimulated with EGF for the appropriate times, lysed, processed, and analyzed to quantify temporal tyrosine phosphorylation profiles. In order to estimate the analytical reproducibility of the method and ensure that quantification was not biased by any particular isoform of the iTRAQ reagent we performed two replicate analyses on this set of samples, but we reversed the iTRAQ labeling scheme. A subset of the results from these replicate analyses is presented in Table 1. Of the 78 tyrosine phosphorylation sites identified from analysis of the first set of cells, 65 were detected and quantified in the analyses of the second set of stimulated cells. The absence of the remaining 13 phosphorylated peptides from the second data set may be due to variation in the efficiency of the immunoprecipitation, but is more likely due to information-dependent acquisition (IDA) based selection of peaks for MS/MS analysis. With this mode of operation, it is not surprising that there is some variation in the identification of peptides between analytical or biological replicates, since the peaks selected for MS/MS will vary from analysis to analysis based on slight shifts in peak selection timing and chromatography. In fact, although 13 tyrosine phosphorylation sites were missed in the analysis of the second set of cells, an additional 26 tyrosine phosphorylation sites were identified with corresponding quantitative temporal phosphorylation profiles

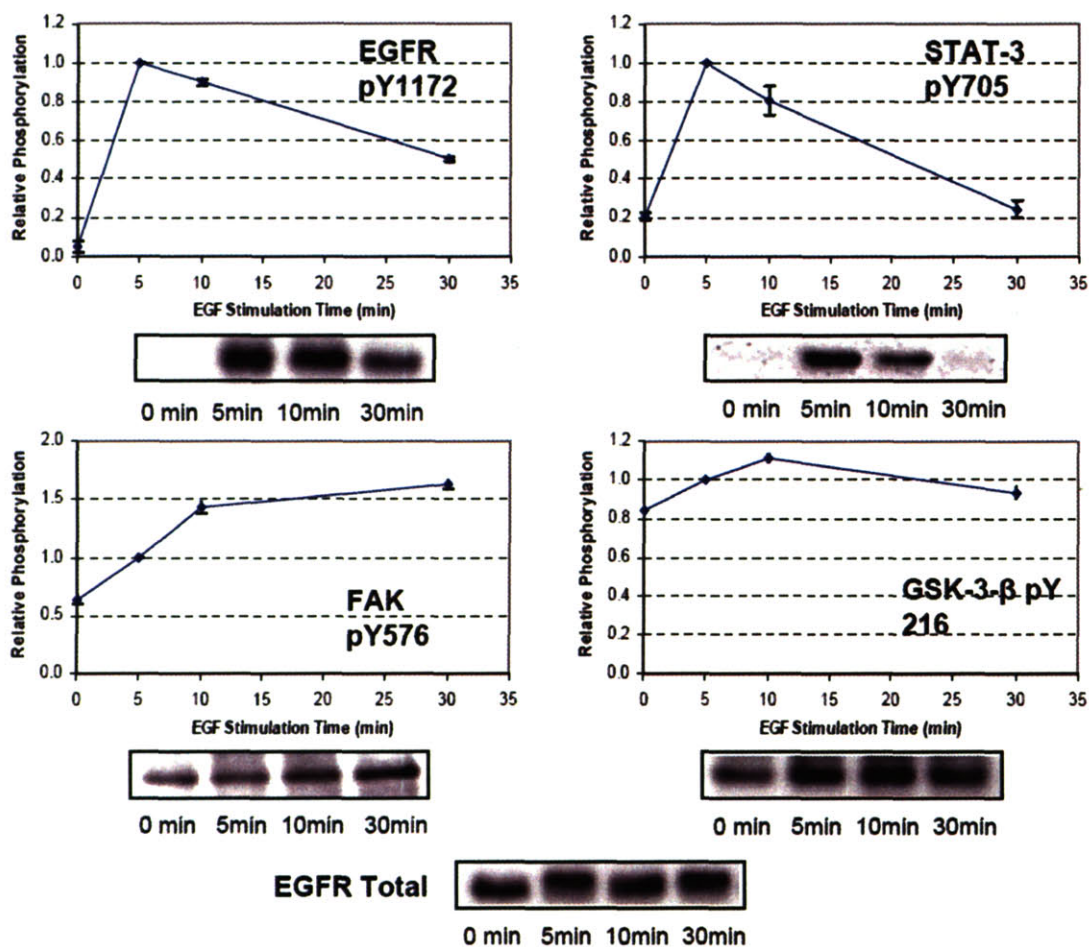


Figure III.3 – Comparison of tyrosine phosphorylation dynamics as measured by mass spectrometry and western blot analysis. Tyrosine phosphorylation profiles determined by mass spectrometry were recapitulated by western blot analysis, further validating the results generated by this approach. Gel loading levels were normalized by probing with an anti-EGFR antibody (bottom).

Because our technology is novel to the biological community we chose to further validate our phosphorylation results by comparing them with those generated by semi-quantitative western blot with phospho-specific antibodies – a methodology widely accepted in biology. To do this comparison we selected four phosphorylation sites identified by mass spectrometry for which we knew high quality antibodies were available. Total cell lysate was loaded on the gel, blotted,

and selected anti-phosphotyrosine antibodies were used to develop the westerns. As demonstrated in Figure III.3, semi-quantitative tyrosine phosphorylation site specific western blots recapitulated the general trend measured by mass spectrometry.

III.3.1.4 Data Variability

In addition to providing validation for the data set from the first set of stimulated cells, analysis of a second set of stimulated cells also provided an estimation of the biological variance between the two sets of cell cultures. Similar stimulation conditions should produce similar results, and as expected, the majority of the temporal phosphorylation profiles are quite similar between the two biological samples. However, some of the profiles show variance in both the level of stimulation (basal (0 minutes) relative to 5 minutes) and the shape of the temporal profile, perhaps reflective of signaling pathways that are potentiated differently between the two sets of cell cultures. Among many possible sources of variation between the cultures slight changes in the confluence of the cells – a high percentage of the variability occurs in proteins related to cell-cell interactions and different cell-cell contacts could activate signaling pathways differently – and changes in the passage number of the cells – the second set of cells were cultured and stimulated 1 month after the first set of cells, which might redound on differential protein expression – might explain the most of differences between the two biological samples. Given the sources of variation, it is not surprising to

see slight alteration of the temporal profiles for some of the phosphorylation sites. Nonetheless, we need to stress the fact that most of the temporal phosphorylation profiles are very similar between the cell cultures which is indicative of the biological reproducibility of our methodology and helps to further validates our results.

III.3.2 Bioinformatic Analysis Reveals Dynamic Modules within the EGFR Signaling Network

To ascribe potential functionality to proteins and phosphorylation sites not previously associated with the EGFR signaling network, we attempted to find similar features in temporal phosphorylation profiles between poorly and well-characterized protein phosphorylation sites. Manually sorting through the data tends to introduce bias and proved to be time-intensive and non-productive. So we resorted to bioinformatic analysis and generated a self-organizing map (SOM) with the dataset from the first set of cells. SOM is a mathematical technique designed to identify underlying patterns in complex data sets; SOMs have been used to analyze gene expression patterns in hematopoietic differentiation, creating biologically relevant clusters and enabling the generation of interesting hypotheses (Tamayo et al., 1999).

To cluster self-similar phosphorylation profiles (modules) within the EGFR signaling network – which might allow us to generate plausible hypotheses on

the role of several phosphorylation sites in the signaling cascade and possible understand their regulation within the network – we applied the SOM algorithm to our quantitative tyrosine phosphorylation profiles. Because, mathematical theory for SOMs has not been refined yet, parameters such as SOM matrix size and distance metrics remain data dependent and need to be empirically determined. We tested several different options for SOM matrix size before settling on a 3 x 3 matrix (Figure III.4A). Smaller matrixes generated clusters that were too complex and difficult to interpret, while larger matrixes separated clusters of phosphorylation sites with similar patterns.

Clustering self-similar phosphorylation profiles revealed interesting modules in the EGFR signaling network and successfully grouped poorly-characterized sites with several well-described proteins in the network. For instance, one of the SOM clusters (Figure III.4B) has the common profile of a large increase in phosphorylation level from basal to 5 minutes followed by slow de-phosphorylation from 5 to 10 and 30 minutes. Included in this cluster is the EGFR pY1172 autophosphorylation site, and two tyrosine phosphorylation sites on SHC, a protein whose PTB (phosphotyrosine-binding) domain binds to EGFR pY1172 (Sakaguchi et al., 1998). In fact, almost all of the phosphorylation sites in this cluster are located on proteins known to interact with EGFR or other receptor tyrosine kinases. For instance, PI3K p85 α has been shown to interact (both directly and indirectly) with both EGFR and PDGFR (Hu et al., 1992). PI3K p85 α pY580 has been ascribed to insulin receptor tyrosine kinase activity (Hayashi et al., 1993), but is most likely due to phosphorylation by EGFR under these

stimulation conditions. Other proteins in this group include c-Cbl, Rho-GEF 5, ACK1, BDP1, Erk1 and hypothetical protein FLJ30532, one of the six proteins not previously associated with the EGFR network. c-Cbl is tyrosine phosphorylated following EGF stimulation and has recently been shown to interact with pY1045 of EGFR following EGF stimulation and prior to endocytosis of the receptor (Grovdal et al., 2004). Tyrosine phosphorylation of activated CDC42 kinase 1 (ACK1) in response to EGF stimulation has been established (Sato et al., 1996). This protein is most likely localized to the receptor through an interaction with CDC42 and Rho-GEF 5; Rho-GEF proteins have been shown to interact directly with tyrosine kinase receptors (Taya et al., 2001). BDP1 is a phosphatase involved in regulation of Gab1, MAPK, and HER2 signaling following EGF stimulation (Gensler et al., 2004), it is likely that phosphorylation of this protein stimulates phosphatase activity in a negative feedback mechanism.

The remaining two proteins in this cluster, Erk1 and hypothetical protein FLJ30532 are not known to associate directly with EGFR or other receptor tyrosine kinases, but from their profiles and grouping in this cluster, the tyrosine phosphorylation sites on these proteins seem to be involved in the immediate early-response to EGF stimulation. In the case of Erk1, although the tyrosine site identified in this study is located in the activation loop, the doubly-phosphorylated, fully active form of the protein was not detected; most likely this peptide was not pulled down by the peptide IP or it was present at a lower level and was below the detection limit of the current approach. The tyrosine kinase that phosphorylates ERK remains unknown, but our results confine its identity to

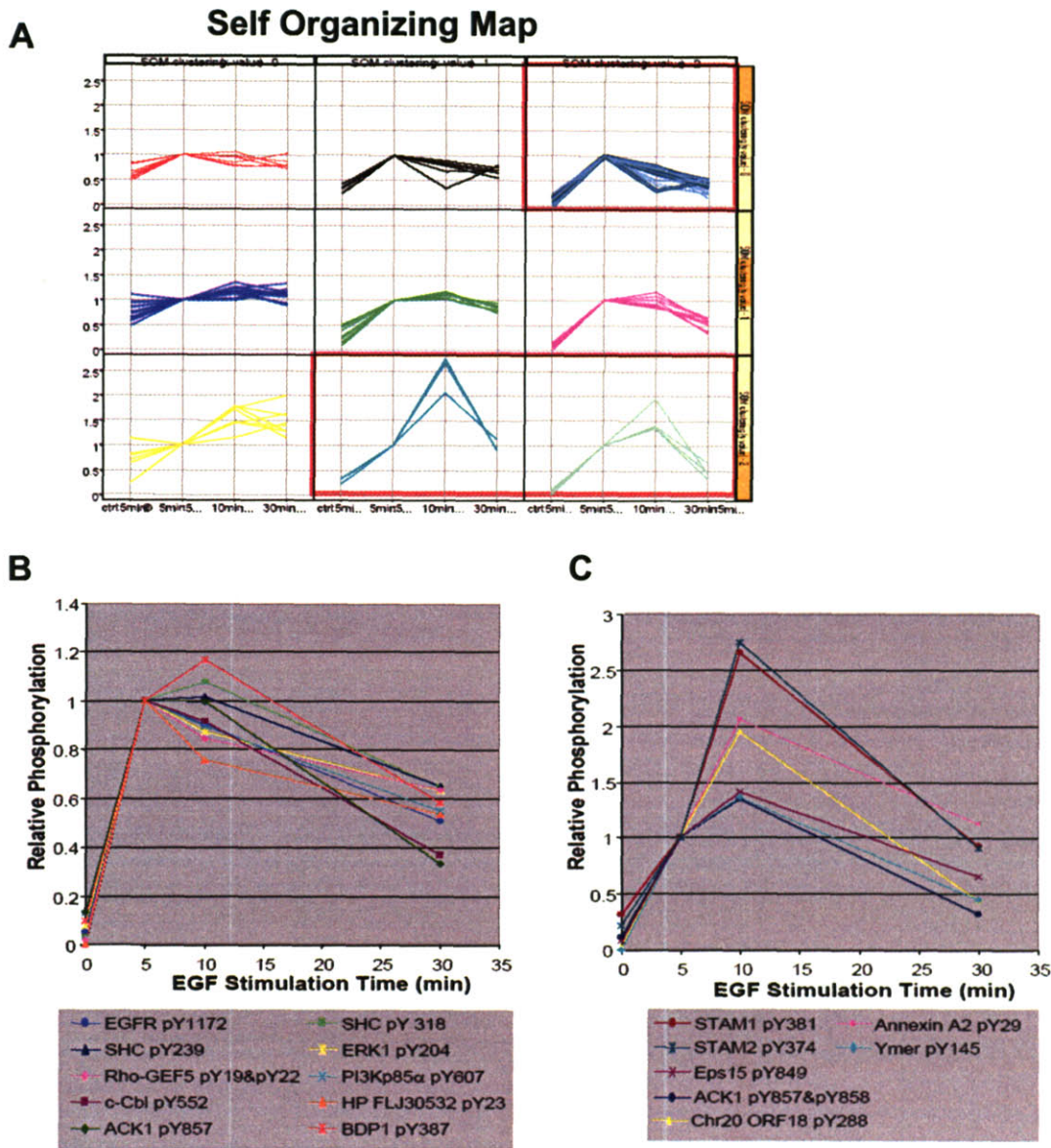


Figure III.4 – Self-organizing maps (SOMs) facilitate the identification of EGFR signaling modules. (A) After loading the mass spectrometry data into Spotfire, a 3 x 3 self-organizing map was generated to facilitate interpretation of the data and to uncover groups of phosphorylation sites that displayed similar patterns. **(B)** A single cluster within the SOM contained many phosphorylation sites involved in the immediate-early response. **(C)** With the exception of the doubly phosphorylated peptide from ACK1, HOIL-1, and Ymer, phosphorylation sites contained in two adjacent clusters are from proteins implicated in regulation of receptor endocytosis and trafficking. Each of these sites was maximally phosphorylated at 10 minutes of EGF stimulation. The level of phosphorylation at 10 and 30 minutes relative to 5 minutes separated the phosphorylation sites into different compartments.

kinases closely downstream of EGFR activation or possibly the receptor itself. However, we require further temporal resolution to refine this claim and biological validation to confirm it. Hypothetical protein FLJ30532 is also known as MARVELD2, named for MARVEL, a recently discovered tetra-spanning transmembrane domain (Sanchez-Pulido et al., 2002). Interestingly, another MARVEL-domain containing protein has recently been implicated in mediating signal transduction from the B-cell receptor to Erk2 (Imamura et al., 2004). Identification of the tyrosine phosphorylation site on MARVELD2 provides a handle by which to investigate the role of this protein in EGFR-mediated signal transduction, while localization to this cluster potentially localizes the topological action of this protein to the immediate-early response to EGF stimulation.

In addition to the immediate-early response module, SOM analysis of the data also uncovered an endocytosis and trafficking module in a meta-cluster of self-similar time-course phosphorylation profiles found in two adjacent compartments of the SOM (Figure III.4C). These phosphorylation sites are all maximally phosphorylated at 10 minutes following EGF stimulation and have separated into two SOM compartments based on phosphorylation level at 10 minutes relative to 5 minutes of stimulation. Proteins within this cluster known to be involved in endocytosis and trafficking included Eps15, STAM1, STAM2, and Annexin A2. EPS15 is necessary for receptor-mediated endocytosis of EGF, and tyrosine phosphorylation of Eps15 following EGF stimulation has been shown (Salcini et al., 1999). Mono-ubiquitination of Eps15 occurs after EGF-induced internalization of the EGF receptor (van Delft et al., 1997), and may be related to

c-Cbl activity, but a phosphorylation site from an alternate E3 ubiquitin-protein ligase, chromosome 20 open reading frame 18 (also known as HOIL-1 (Yamanaka et al., 2003)) is localized to this cluster and may be responsible for further ubiquitination within the endosomes. Involvement of HOIL-1 in the EGFR pathway has not been characterized, but its biological function fits in well with endocytosis and trafficking. STAM1 and STAM2 phosphorylation sites are also located in these clusters, these proteins bind to ubiquitinated proteins via their VHS domains and ubiquitin-interacting motifs (UIM) in early endosomes and are involved in endosomal trafficking (Mizuno et al., 2003). The N-terminus of Annexin A2 is tyrosine-phosphorylated following growth factor stimulation and may play a role in receptor trafficking; recently it has been shown that phosphorylation is blocked when receptor internalization is inhibited (Biener et al., 1996). Based on the clustering of these phosphorylation sites, it is possible that phosphorylation of these sites occurs following EGFR ubiquitination by Cbl and endocytosis via clathrin-coated early endosomes. If so, it may be that Ymer, a protein recently found to be modulated by EGF stimulation (Blagoev et al., 2004) and located in these clusters, may be localized to the plasma membrane and early endosomes. If Ymer follows the role of the other proteins within this cluster, tyrosine phosphorylation of this protein may regulate endosomal trafficking of the receptor. Identification of this regulated site will enable further investigation to determine the functional role of phosphorylation of this protein in EGFR signaling. Along with this line of argument a recent publication by Konishi et al. (Tashiro et al., 2006) shows that Ymer co-localizes and directly interacts

with EGFR on the plasma membrane and when over-expressed inhibits receptor endocytosis, degradation and post-translational modification arguing that Ymer plays a role in regulating EGFR activity and sub-cellular localization.

III.4 DISCUSSION

Quantitative, time-resolved analysis of phosphorylated-tyrosine mediated signaling cascades downstream of activated receptor tyrosine kinases will be critical in developing a better understanding of cell signaling and the molecular mechanisms underlying a variety of disease states. To this end, the methodology that we have developed has enabled the quantification of the temporal phosphorylation profile of specific tyrosine phosphorylation sites following EGF stimulation. This methodology is not limited to EGFR signaling, but is applicable to the investigation of a broad variety of other tyrosine-mediated signal transduction pathways from cell culture and notably *in vivo* experiments. In fact, because stable-isotope labeling of peptides with the iTRAQ reagent occurs after cell lysis and tryptic digestion but prior to immunoprecipitation, quantification is not limited to cell culture-derived samples as is SILAC (Ong et al., 2002) and variations in immunoprecipitation do not affect quantification when comparing four samples or less. For larger sets as those that will be discussed in chapters III and IV normalization between experiments is required to take into account variability in immunoprecipitation and LC-MS-MS/MS runs. However, arguably the highest success of our methodology relies in our mass spectrometry sensitivity which allows us to use only several hundred micrograms of protein per sample per analysis, hence this method should be readily amenable to the

investigation of signal transduction in samples with limited quantities of protein, such as clinical tumor tissues.

It is worth noting that Blagoev et al. (Blagoev et al., 2004) have published an alternative method combining SILAC and immunoprecipitation of tyrosine phosphorylated proteins, thereby enabling the quantification of temporal involvement in the EGFR signaling network at the protein level. Although the two methods appear quite similar, there are important advantages and disadvantages of each that should be noted. Whereas our method provides identification of specific protein phosphorylation sites and quantitative temporal phosphorylation profiles for each of these sites; the Blagoev et al. (Blagoev et al., 2004) method monitors tyrosine phosphorylation (or changes in association with tyrosine phosphorylated proteins) at the protein level, without site specification. Our methodology advantage is that it provides site-specific monitoring of protein phosphorylation and hence more explicit detail regarding the regulation of proteins within the network, but its disadvantage is that it precludes analysis of potential other, non-phosphorylated proteins associated (and therefore co-immunoprecipitated) with tyrosine phosphorylated proteins that might also play a crucial role in the signaling pathways as could be adaptor proteins or important members of the MAPK pathway as Grb2 and SOS. Both methods are limited by the amenability of peptides (and therefore proteins) to be identified by LC/MS/MS analysis, as many peptides might be immunoprecipitated, but go undetected in the mass spectrometry runs. This limitation is more severe for peptide-immunoprecipitation where there are fewer peptides with which to identify each

protein. Our current method utilizes only a single anti-phosphotyrosine antibody, such that the results may be constrained by the bias associated with this particular antibody, a problem that is addressed in chapters III and IV. A second antibody related problem is quality variability from lot to lot for any given antibody, which precludes from working with a stable supplier (we started this work using anti-phosphotyrosine antibody 4G10 from Upstate Technologies, but had to change to anti-phosphotyrosine antibody P99 from Santa-Cruz Biotechnologies) Despite differences between the two methods, there is significant consistency at the protein level between the Blagoev et al. dataset and our dataset, with our data providing additional site-specific information. Differences in the temporal phosphorylation profiles between the two studies are most likely caused by multiple tyrosine phosphorylation sites on a given protein, but may also be due to the different quantification methods (SILAC vs. iTRAQ).

Another study utilizing anti-phosphotyrosine protein immunoprecipitation, stable isotope (ICAT) labeling and mass spectrometry to study EGFR signaling was recently published (Thelemann et al., 2005). Similar to Blagoev et al. (Blagoev et al., 2004), Thelemann et al. (Thelemann et al., 2005) monitored tyrosine phosphorylation (or association with tyrosine phosphorylated proteins) at the protein level, without site specification, although many phosphorylation sites were identified without stable-isotope quantification in a separate section of the manuscript. It is difficult to directly compare our data to that reported by Thelemann et al. (Thelemann et al., 2005) as the total number of proteins quantified in the Thelemann et al. (Thelemann et al., 2005) paper was

significantly fewer than contained in either our dataset or the dataset reported by Blagoev et al. (Blagoev et al., 2004).

In this chapter I have described a mass spectrometry based technology which was developed by a collaborative effort within the White lab and which is the base of chapters III and IV of this thesis. This methodology provides a reliable analytical method enabling the quantification of time-resolved tyrosine phosphorylation profiles with site specific resolution and allowing the estimation of the statistical significance on changes in phosphorylation levels between the different time points following stimulation.

Data generated by this method and described in this chapter self-organizes into clusters of phosphorylation sites that recapitulate and extend biological findings in the literature. Generation of a SOM has enabled the identification of modules in the EGFR signaling network. Based on the presence of particular proteins within these modules, we can generate additional hypotheses aimed at defining the potential function of tyrosine phosphorylation on several proteins and/or phosphorylation sites not previously implicated in EGFR signaling.

Subsequent application of this method to other signaling cascades has already considerably improved our knowledge on TCR and insulin activated pathways (Kim & White, 2006; Schmelzle et al., 2006) it is been currently use in the White lab to study aberrant signaling downstream of constitutively active and over-expressed tyrosine kinases. We expect that these studies may reveal mechanisms of pathogenesis in these systems, and may provide additional

targets for novel therapeutics. In particular chapter III of this thesis will describe how application of this technique to understand the role of HER2 over-expression in EGFR family activated signaling pathways relate to cell proliferation and migration, two phenotypes closely related to cancer progression.

III.5 REFERENCES

Balsalobre A, Marcacci L, Schibler U (2000) Multiple signaling pathways elicit circadian gene expression in cultured Rat-1 fibroblasts. *Curr Biol* **10**: 1291-1294.

Band V, Sager R (1989) Distinctive traits of normal and tumor-derived human mammary epithelial cells expressed in a medium that supports long-term growth of both cell types. *Proc Natl Acad Sci U S A* **86**: 1249-1253.

Biener Y, Feinstein R, Mayak M, Kaburagi Y, Kadowaki T, Zick Y (1996) Annexin II is a novel player in insulin signal transduction. Possible association between annexin II phosphorylation and insulin receptor internalization. *J Biol Chem* **271**: 29489-29496.

Blagoev B, Ong SE, Kratchmarova I, Mann M (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat Biotechnol* **22**: 1139-1145.

Di Stefano P, Cabodi S, Boeri Erba E, Margaria V, Bergatto E, Giuffrida MG, Silengo L, Tarone G, Turco E, Defilippi P (2004) P130Cas-associated protein (p140Cap) as a new tyrosine-phosphorylated protein involved in cell spreading. *Mol Biol Cell* **15**: 787-800.

Ficarro SB, McClelland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM (2002) Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat Biotechnol* **20**: 301-305.

Gensler M, Buschbeck M, Ullrich A (2004) Negative regulation of HER2 signaling by the PEST-type protein-tyrosine phosphatase BDP1. *J Biol Chem* **279**: 12110-12116.

Grovdal LM, Stang E, Sorkin A, Madshus IH (2004) Direct interaction of Cbl with pTyr 1045 of the EGF receptor (EGFR) is required to sort the EGFR to lysosomes for degradation. *Exp Cell Res* **300**: 388-395.

Hayashi H, Nishioka Y, Kamohara S, Kanai F, Ishii K, Fukui Y, Shibasaki F, Takenawa T, Kido H, Katsunuma N (1993) The alpha-type 85-kDa subunit of phosphatidylinositol 3-kinase is phosphorylated at tyrosines 368, 580, and 607 by the insulin receptor. *J Biol Chem* **268**: 7107-7117.

Hendriks BS, Opresko LK, Wiley HS, Lauffenburger D (2003) Coregulation of epidermal growth factor receptor/human epidermal growth factor receptor 2 (HER2) levels and locations: quantitative analysis of HER2 overexpression effects. *Cancer Res* **63**: 1130-1137.

Hu P, Margolis B, Skolnik EY, Lammers R, Ullrich A, Schlessinger J (1992) Interaction of phosphatidylinositol 3-kinase-associated p85 with epidermal growth factor and platelet-derived growth factor receptors. *Mol Cell Biol* **12**: 981-990.

Imamura Y, Katahira T, Kitamura D (2004) Identification and characterization of a novel BASH N terminus-associated protein, BNAS2. *J Biol Chem* **279**: 26425-26432.

Kim JE, White FM (2006) Quantitative analysis of phosphotyrosine signaling networks triggered by CD3 and CD28 costimulation in Jurkat cells. *J Immunol* **176**: 2833-2843.

Kramer A, Yang FC, Snodgrass P, Li X, Scammell TE, Davis FC, Weitz CJ (2003) Regulation of daily locomotor activity and sleep by hypothalamic EGF receptor signalling. *Novartis Found Symp* **253**: 250-262; discussion 102-259, 263-256, 281-254.

Mizuno E, Kawahata K, Kato M, Kitamura N, Komada M (2003) STAM proteins bind ubiquitinated proteins on the early endosome via the VHS domain and ubiquitin-interacting motif. *Mol Biol Cell* **14**: 3675-3689.

Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**: 376-386.

Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlett-Jones M, He F, Jacobson A, Pappin DJ (2004b) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **3**: 1154-1169.

Rush J, Moritz A, Lee KA, Guo A, Goss VL, Spek EJ, Zhang H, Zha XM, Polakiewicz RD, Comb MJ (2004) Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat Biotechnol*.

Sakaguchi K, Okabayashi Y, Kido Y, Kimura S, Matsumura Y, Inushima K, Kasuga M (1998) Shc phosphotyrosine-binding domain dominantly interacts with epidermal growth factor receptors and mediates Ras activation in intact cells. *Mol Endocrinol* **12**: 536-543.

Salcini AE, Chen H, Iannolo G, De Camilli P, Di Fiore PP (1999) Epidermal growth factor pathway substrate 15, Eps15. *Int J Biochem Cell Biol* **31**: 805-809.

Sanchez-Pulido L, Martin-Belmonte F, Valencia A, Alonso MA (2002) MARVEL: a conserved domain involved in membrane apposition events. *Trends Biochem Sci* **27**: 599-601.

Satoh T, Kato J, Nishida K, Kaziro Y (1996) Tyrosine phosphorylation of ACK in response to temperature shift-down, hyperosmotic shock, and epidermal growth factor stimulation. *FEBS Lett* **386**: 230-234.

Schmelzle K, Kane S, Gridley S, Lienhard GE, White FM (2006) Temporal dynamics of tyrosine phosphorylation in insulin signaling. *Diabetes* **55**: 2171-2179.

Stampfer MR, Bartley JC (1985) Induction of transformation and continuous cell lines from normal human mammary epithelial cells after exposure to benzo[a]pyrene. *Proc Natl Acad Sci U S A* **82**: 2394-2398.

Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci U S A* **96**: 2907-2912.

Tang CK, Perez C, Grunt T, Waibel C, Cho C, Lupu R (1996) Involvement of heregulin-beta2 in the acquisition of the hormone-independent phenotype of breast cancer cells. *Cancer Res* **56**: 3350-3358.

Tashiro K, Konishi H, Sano E, Nabeshi H, Yamauchi E, Taniguchi H (2006) Suppression of the ligand-mediated downregulation of epidermal growth factor

receptor by Ymer, a novel tyrosine phosphorylated and ubiquitinated protein. *J Biol Chem* **27**: 27.

Taya S, Inagaki N, Sengiku H, Makino H, Iwamatsu A, Urakawa I, Nagao K, Kataoka S, Kaibuchi K (2001) Direct interaction of insulin-like growth factor-1 receptor with leukemia-associated RhoGEF. *J Cell Biol* **155**: 809-820.

Thelemann A, Petti F, Griffin G, Iwata K, Hunt T, Settinaro T, Fenyo D, Gibson N, Haley JD (2005) Phosphotyrosine signaling networks in epidermal growth factor receptor overexpressing squamous carcinoma cells. *Mol Cell Proteomics* **4**: 356-376.

van Delft S, Govers R, Strous GJ, Verkleij AJ, van Bergen en Henegouwen PM (1997) Epidermal growth factor induces ubiquitination of Eps15. *J Biol Chem* **272**: 14013-14016.

Yamanaka K, Ishikawa H, Megumi Y, Tokunaga F, Kanie M, Rouault TA, Morishima I, Minato N, Ishimori K, Iwai K (2003) Identification of the ubiquitin-protein ligase that recognizes oxidized IRP2. *Nat Cell Biol* **5**: 336-340.

Zhang Y, Wolf-Yadlin A, Ross PL, Pappin DJ, Rush J, Lauffenburger DA, White FM (2005b) Time-resolved Mass Spectrometry of Tyrosine Phosphorylation Sites in the Epidermal Growth Factor Receptor Signaling Network Reveals Dynamic Modules. *Mol Cell Proteomics* **4**: 1240-1250.

**IV Quantitative Phosphoproteomic Analysis of HER2-
overexpression effects on Cell Signaling Networks
governing proliferation and migration**

IV.1 SUMMARY

Although HER2 over-expression is implicated in tumor progression for a variety of cancer types, how it disregulates signaling networks governing cell behavioral functions is poorly understood. In this chapter I will portray the application of our mass spectrometry technology based technology – described in chapter III – to address this problem.

In this chapter we extended our technology to analyze 16 different biological samples. This allowed us to obtain quantitative mass spectrometry measurements of the dynamic effect of HER2 over-expression on phosphotyrosine signaling in human mammary epithelial cells (HMECs) stimulated by EGF or HRG. We further analyzed HMEC proliferation and migration under the same set of experimental conditions.

We used three different, but complementary computational approaches to analyze the data. Hierarchical clustering (static) and self-organizing map (dynamic) analysis of the resulting phosphoproteomic dataset permitted elucidation of network modules differentially regulated in HER2-overexpressing cells in comparison with parental cells for EGF and HRG treatment. Partial least-squares regression analysis of the same dataset identified quantitative combinations of signals within the networks that strongly correlated with our cell proliferation and migration measured measurements. Combining these modeling approaches enabled association of EGFR-family dimerization to activation of

specific phosphorylation sites which appear to most critically regulate proliferation and/or migration.

Overall, through the work that I will present in this chapter we were able to characterize quantitative relationships between ligand stimulation, cell signaling, and cellular response that explain how HER2 over-expression can alter network signals and downstream cell behavior.

The work presented in this chapter has been accepted for publication by Molecular Systems Biology (MSB) and it represents a collaborative effort between the White lab and Lauffenburger lab. In particular Self Organizing Maps analysis of the data was done by Sampsa Hautaniemi in the Lauffenburger lab (currently a Professor in The University of Helsinki). Proliferation, migration and partial least square regression models were done by Neil Kumar in the Lauffenburger lab.

IV.2 Materials and Methods

IV.2.1 Cell Culture and Stimulation

184A1 HMECs (human mammary epithelial cells (Stampfer & Bartley, 1985) (HMEC Parental) were a kind gift from Martha Stampfer (Lawrence Berkeley Laboratory, Berkeley CA) via Steve Wiley (Pacific Northwest National Laboratory, Richland WA) and were maintained in DFCI-1 medium supplemented with 12.5 ng/ml EGF, as in (Band & Sager, 1989). 184A1 HMECs clone 24H (human mammary epithelial cells over-expressing HER2 30 fold (Hendriks et al., 2003a)) (HMEC 24H) were a kind gift from Steve Wiley (Pacific Northwest National Laboratories, Richland WA) and were maintained in DFCI-1 medium supplemented with 12.5 ng/ml EGF and 150 µg/ml of Geneticin. Cells were washed with PBS and incubated for 12 hours in serum free media (DFCI-1 without EGF, bovine pituitary extract, or fetal bovine serum) after 80% confluence was reached in 15 cm plates (~10⁷ cells). Synchronized cells were washed with PBS after removal of media. Cells were then stimulated with 100 ng/ml EGF or 80 ng/ml HRG in serum free media for 5, 10 or 30 minutes, or left untreated with serum free media for 5 min as control before lysing.

IV.2.2 Protein Quantification

Protein concentration of cell lysates was determined using the micro bicinchoninic acid assay (micro-BSA – Pierce) according to the manufacturer's protocol.

IV.2.3 Cell Lysis, Protein Digestion, Peptide Fractionation and iTRAQ Labeling

Cells were lysed with 8M urea / 1mM sodium orthovanadate after EGF or HRG stimulation. Proteins were digested with trypsin after DTT reduction and iodoacetamide alkylation. Tryptic peptides were desalted and fractionated on a C18 Sep-Pak Plus cartridge (Waters), and the 25% MeCN fraction was divided into 10 equivalent aliquots which were lyophilized to dryness. One aliquot of sample from each condition was labeled with one tube of iTRAQ reagent (Applied Biosystems, CA) (following manufacturer's directions). Samples labeled with four different isoforms of the iTRAQ reagent were combined, dried completely, and saved at -80°C. This process was repeated to generate five duplicate sets of samples: four time-course samples (0, 5, 10, 30 min) with 100 ng/ml EGF or 80 ng/ml HRG stimulation in either HMEC or 24H cells, and one 5 min mix sample that consisted of the samples stimulated for 5 min for each of the

stimulation conditions in the order of HMEC/HRG, HMEC/EGF, 24H/HRG, 24H/EGF.

IV.2.4 Peptide Immunoprecipitation

10 µg of protein G Plus-agarose beads (Calbiochem) were incubated with 3.5 µg of each anti-phosphotyrosine antibody (PY99(Santa Cruz), 4G10(Upstate) and pTyr100 (Cell Signaling Technology)) in 200 µl of IP buffer for 8 hr at 4°C. The beads were rinsed once with 400 µl of IP buffer at 4°C. iTRAQ labeled sample were dissolved in 150 µl of IP buffer (100 mM Tris, 100 mM NaCl, 1% NP-40, pH 7.4) and 300ul of water. After pH was adjusted to 7.4 with 0.5 M Tris buffer pH 8.5, the sample was mixed with protein G Plus-agarose beads, and was incubated overnight at 4°C. The protein G Plus-agarose beads were spun down for 5 minutes at 6000 rpm and the supernatant was separated and saved. Antibody-bound beads were washed once with 400 µl IP buffer for 10 minutes and twice with rinse buffer (100 mM Tris, 100 mM NaCl, pH 7.4) for 5 minutes at 4°C. Phosphotyrosine containing peptides were eluted from antibody with 60 µl of 100 mM glycine pH 2.5 for 30 min at room temperature.

IV.2.5 IMAC and Mass Spectrometry

Phosphopeptide enrichment on IMAC was performed as described (Zhang et al., 2005c). Peptides retained on the IMAC column were eluted to a C₁₈ capillary precolumn (100 µm ID, 360 µm OD) with 50 µl of 250 mM Na₂HPO₄, pH 8.0. After 10 min rinse (0.1% HOAc), the precolumn was connected to an analytical capillary column with an integrated electrospray tip (~1 µm orifice). Peptides were eluted using a 100 min gradient with solvent A (H₂O/HOAc, 99/1 vol/vol) and B (H₂O/MeCN/HOAc, 29/70/1 vol/vol): 10 min from 0% to 15% B, 75 min from 15% to 40% B, 15 min from 40% to 70% B. Eluted peptides were directly electrosprayed into a quadrupole time-of-flight mass spectrometer (QSTAR XL Pro, Applied Biosystems). MS/MS spectra of four or five most intense peaks with 2 to 5 charges in the MS scan were automatically acquired in information-dependent acquisition with previously selected peaks excluded for 25 seconds.

IV.2.6 Phosphopeptide Sequencing, Data Clustering and Analysis

MS/MS spectra were extracted and searched against human protein database (NCBI) using ProQuant (Applied Biosystems) as described. An interrogator database was generated by predigesting the human protein database with trypsin and allowing one miscleavage and up to six modifications

on a single peptide (phosphotyrosine ≤ 2 , phosphoserine ≤ 1 , phosphothreonine ≤ 1 , iTRAQ-lysine ≤ 4 , and iTRAQ-tyrosine ≤ 4). Mass tolerance was set to 2.3 amu for precursor ions and 0.15 amu for fragment ions and grouping of MS/MS spectra from different cycles was set to zero. Phosphotyrosine-containing peptides identified by ProQuant were manually validated. ProQuant quantification results were corrected by removing contaminant signals near iTRAQ tag peaks. Data was further corrected with values generated from the peak areas of non-phosphorylated peptides to account for possible variations in the starting amounts of sample for each time point. All the data from each analysis was normalized by the 115 peak area values, which correspond to the 5 min sample in each time-course analysis or HMEC-EGF sample in the four-condition 5 min mix analysis.

IV.2.7 ELISA protocol for ErbB3 receptor quantification

A black 96-well Nunc MaxiSorp plate was coated with 50 μ L of capture antibody (4 μ g/mL in PBS) overnight at room temperature and washed three times with PBS, 0.05% Tween-20 on a Bio-Tek plate washer. The unreacted surface of the plate was blocked with 300 μ L of 2% BSA in PBS for 1 hr at room temperature. After washing, 50 μ L of each sample or the recombinant standard were added to the appropriate wells. Cell line lysates were mixed 1:1 with 2% BSA and added to the plate for 2 hr incubation at 4°C. Both standards and

samples were done in duplicate. After washing, 60 μ L of detection antibody (2 mg/mL diluted in 2 %BSA, PBS, 0.1% Tween-20) was added to all the wells and incubated for 1 hr at room temperature. The plate was washed and 60 μ L of Streptavidin-HRP (diluted 1:200 in 2 %BSA, PBS, 0.1% Tween-20) was added for 30 min at room temperature. After a final wash, 60 μ L of chemiluminescent substrate (SuperSignal ELISA Pico from Pierce) was added to all wells and the plate was read immediately on a Fusion plate reader (Perkin Elmer) with a 1 sec reading time per well. The data was exported to Excel, replicates were averaged and the standard dilutions were fit to a linear curve.

IV.2.8 Hierarchical Clustering

After normalization to the 5 minute HMEC-EGF sample, hierarchical clustering of the four-condition 5min Mix analysis was done using Spotfire™ using the Euclidean distance and unweighted average settings.

IV.2.9 Self Organizing Maps

The SOM non-linearly transforms high-dimensional input data into a lower dimensional display that consists of several map units. Each map unit consists of

a reference vector whose dimension is the same as dimension of the input pattern (Kohonen, 2001). The clustering is done in two steps. In the training phase the SOM algorithm sequentially computes distances between each input pattern and all reference vectors. The map unit consisting of reference vector most close to the input pattern is declared as the winner map unit and is updated with maximal strength, while topologically close map units are updated with gradually decreasing strength. This process is repeated many times, which eventually leads to self-organization. After the training phase each peptide is clustered to the map unit whose reference vector is the most similar to the input pattern.

The parameters for the SOM used in this study were as follows. Topology of the map was chosen to be sheet, distance measure was correlation and the number of map units was chosen to be $5 \cdot \sqrt{p}$, where p is the number of input vectors as suggested in (Vesanto et al., 2000). We used the batch learning algorithm and the neighborhood function was chosen to be Gaussian with the parameters given in (Vesanto et al., 2000). The SOM analysis was performed in MATLAB with the publicly available SOM Toolbox (Vesanto et al., 2000). We used the U-matrix (unified distance matrix) method to identify a group of map units that comprehend a cluster (Ultsch & Siemon, 1989). The U-matrix illustrates Euclidean distances between the map units with colors. Adjacent map units colored with blue shades constitute a cluster (valleys) and the clusters are separated by red and yellow colors that represent high distance between two map units (mountains). For each cluster we computed statistical significance with

the permutation test based method given in (Hautaniemi et al., 2003). Briefly, first we computed the correlation distances between all the peptide profiles in a cluster. If the profiles correlate, their distance is zero and for a pair of negatively correlating profiles the distance value is two. Then we computed the mean of these pair-wise comparisons. This was followed by choosing randomly as many profiles as there were in the original cluster and computing all the pair-wise distances, and taking the mean. If the mean distance of a randomly chosen set is less or equal than the original we added one to counter. For each cluster we created 5,000 samples and the final *p*-value is computed by dividing the counter by 5,000. High values mean that the original cluster may be due to chance.

IV.2.10 Proliferation

Human mammary epithelial cells were plated on 24 well tissue culture plastic plates ($\sim 1 \times 10^4$ cells/cm²) and grown for 24 hours to ~ 60 -70% confluence. The medium was then removed and cells were serum-starved as previously described for 12-16 hours. Starved cells were then treated with new serum-free media, serum-free media containing EGF (100 ng/ml), or serum-free media containing HRG (80 ng/ml). The cells were grown for 15 hours at which time [³H] thymidine (10 μ Ci/ml) was added. After 10 hours, the cells were washed with cold PBS and treated with trichloroacetic acid (5% w/v) at 4°C. The resulting

precipitate was dissolved in NaOH (0.5 N) and quantified using liquid scintillation counting.

IV.2.11 Migration

HMECs were seeded in 96-well tissue culture plastic plates (Packard View Plate Black, Ref. 6005225) at confluence ($\sim 50,000$ cells/cm²) and allowed to adhere for 4-6 hours. Media was then removed and cells were serum starved for 12-16 hours as previously described. Starved cells were treated with 5-chloromethylfluorescein diacetate (CMFDA, 9 μ M in serum-free media) for 30 minutes. CMTFDA containing media was removed and cells were then treated with new serum-free media, serum-free media containing EGF (100 ng/ml), or serum-free media containing HRG (80 ng/ml). A wound width ~ 700 μ m was scraped in each well and cell movement imaged every 15 minutes for 12 hours using Cellomics KineticScan. Kinetics of wound closure were quantified using an in-house analysis software that calculated the wound area at each time point normalized by the initial wound area. A time averaging algorithm was used to average wound closure in four wells of equal treatment into a single trajectory at 30 minute intervals. The trajectories shown here have been additionally normalized to their 2 hour time points for the purpose of comparing treatments. The wound closure data used in the described PLSR model is the linearly fitted slope (2 - 5.5 hours) of each trajectory shown.

IV.2.12 Partial Least Squares Regression

Computational analysis was performed using the SIMCA-P 10.0 (Umetrics) software suite as detailed elsewhere (Gaudet et al., 2005; Janes et al., 2004). The software uses the non-linear iterative partial least-squares algorithm to perform decompositions and regressions. The model was evaluated for goodness of fit (R^2Y), goodness of prediction (Q^2), and was validated against over-fitting via response permutation (response matrix was randomly permuted 50 times and Q^2 values were obtained for each run). All matrices were mean centered and unit variance scaled (z-score normalized) prior to partial least squares analysis.

Goodness of prediction (Q^2) was evaluated using a leave-one-out cross validation approach (Eriksson et al., 2001). Briefly, cross validation is performed by omitting an observation from the model development and then using the model to predict the Y-block values for the withheld observation. The procedure is repeated until every observation has been kept out exactly once. The prediction error sum of squares (PRESS) is then calculated as follows:

$$PRESS = \sum_i \sum_m Y_{i,m}^{measured} - Y_{i,m}^{predicted} \quad (1)$$

Q^2 is then calculated as:

$$Q^2 = 1.0 - \prod_{a=1}^A (PRESS/SS)_a \quad (2)$$

Where a refers to each individual principal component in the model and SS is the sum of squares explained by principal component a . The variable importance for projection (VIP) value for each variable (k) was calculated according to the following formula:

$$VIP_k = \sqrt{\frac{K_T \sum_{a=1}^A w_{a,k}^2 SS_a}{\sum_{a=1}^A SS_a}} \quad (3)$$

Where K_T is the total number of variables and the rest of the variables are as described above.

IV.3 RESULTS

IV.3.1 Phosphotyrosine mass spectrometry of parental and HER2-overexpressing HMECs in response to EGF and HRG treatment

We have undertaken a quantitative investigation of the effects of HER2-overexpression in a human mammary epithelial cell (HMEC) line (184A1 (Stampfer & Bartley, 1985)). We compared the parental line (denoted "P"), which exhibits approximately 200,000 EGFR per cell, 20,000 HER2 per cell, and 20,000 HER3 per cell, to a stable retrovirally-transduced clone (denoted "24H") that expresses HER2 at 600,000 per cell while maintaining constant EGFR levels and increased HER3 levels to about 30,000 per cell. Previously performed combined experimental and modeling studies on this cell line by Bart Hendriks in the Lauffenburger lab has enabled quantitative estimation of the various receptor dimer species under various treatment conditions (Hendriks et al., 2003a; Hendriks et al., 2005). From that work, we expect that 100 ng/ml EGF treatment should result in high levels of EGFR homodimers in the P HMECs (~80,000), but lower numbers of EGFR-HER2 heterodimers (~20,000). By comparison, treating 24H HMECs with this same concentration of EGF should drive a large increase in heterodimers (to ~150,000) and a significant decrease in EGFR homodimers (~10,000). 80 ng/ml HRG treatment should yield ~15,000 HER2-HER3 heterodimers in the P HMECs and in the 24 H HMECs this number should increase to ~25,000. Interestingly, under both ligand treatments, 24H HMECs

are expected to have large numbers of HER2 homodimers (>~200,000), some of which may be activated through basal auto-phosphorylation. In this chapter I will describe our efforts to explore the consequences of these changing dimerization states on intracellular signaling and the subsequent control of cell proliferation and migration. We have acquired mass spectrometry data describing the temporal dynamics (0, 5, 10, and 30 minute stimulation) of tyrosine phosphorylation in the 184A1 cell system under four conditions: P HMECs stimulated with EGF, P HMECs stimulated with HRG, 24H cells HMECs stimulated with EGF, and 24H HMECs stimulated with HRG (see Figure IV.1).

As a result of these analyses, 332 phosphorylated peptides from 175 proteins were identified, including 289 singly (tyrosine) phosphorylated peptides, 42 doubly phosphorylated peptides (21 tyrosine/tyrosine, 18 serine/tyrosine, and 3 threonine/tyrosine), and one triply phosphorylated peptide (tyrosine/tyrosine/tyrosine). A total of 20 phosphorylation sites were identified on EGFR, HER2 and HER3, including 9 tyrosine and 2 serine sites on EGFR, 8 tyrosine phosphorylation sites on HER2, and 1 tyrosine phosphorylation site on HER3 (Figure IV.1, B-D). Of the 20 phosphorylation sites on EGFR family members, Y1114 on EGFR and Y1005 and Y1127 on HER2 represent novel sites which have not been previously described in the literature, although synthetic phosphopeptides containing EGFR Y1114 and HER2 Y1005 have been shown to bind SHC (Schulze et al., 2005) and mutation of EGFR Y1114 has been shown to block SOCS recruitment to the receptor, regulating STAT activation (Xia et al., 2002). Downstream of the receptors, quantitative data was

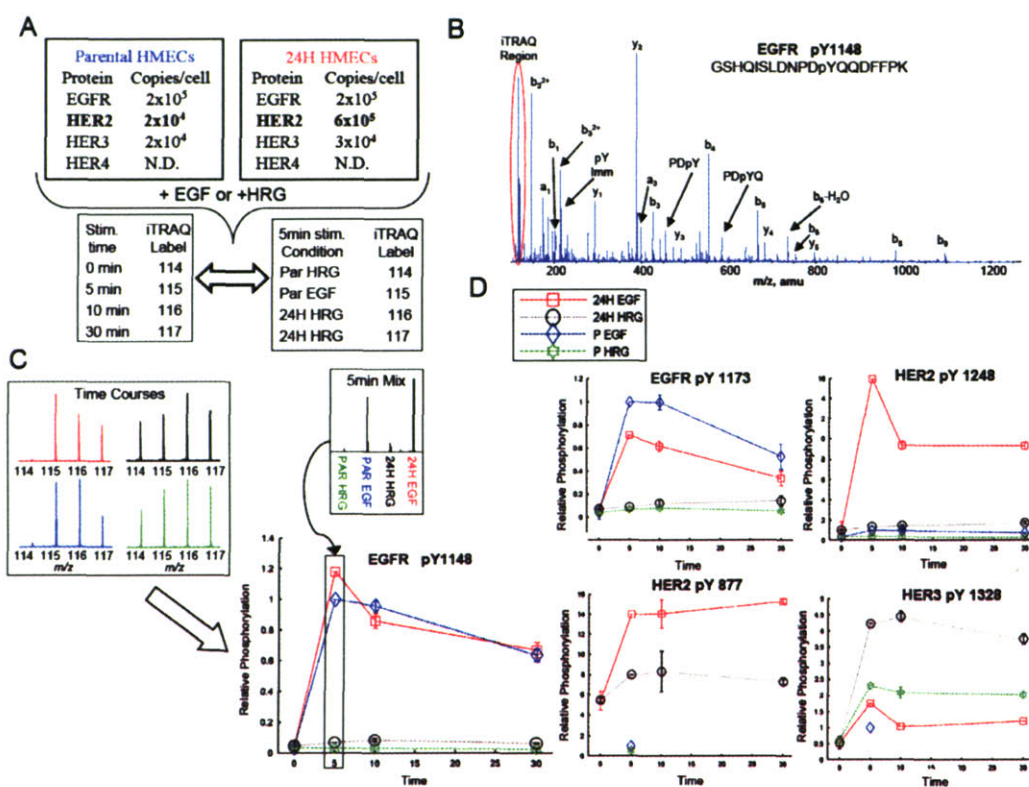


Figure IV.1 – Data acquisition and quantification scheme with example data. (A) 184A1 parental HMEC cells and 24H HMEC cells were stimulated with either EGF or HRG for 0, 5, 10 or 30 minutes. For each stimulation condition, following cell lysis and tryptic digestion, peptides from each time point were labeled with iTRAQ, mixed, and analyzed by anti-phosphotyrosine peptide immunoprecipitation and IMAC-LC-MS/MS, producing temporal tyrosine phosphorylation profiles for hundreds of peptides. To normalize temporal profiles from each condition, tryptic peptides from the 5 minute stimulation time point for each condition were labeled with iTRAQ, mixed, and analyzed by anti-phosphotyrosine peptide immunoprecipitation and IMAC-LC-MS/MS. (B) For each peptide, y - and b -type ions in the MS/MS spectrum provided sequence identification and site assignment, while the iTRAQ marker ion region (highlighted in the red oval) provided quantification. (C) iTRAQ marker ion profiles representing temporal phosphorylation profiles for EGFR pY1148 under all four conditions (top left) as well as the relative amount of phosphorylation following a 5 minute stimulation for each condition (top right). The final plot of temporal response under multiple stimulation conditions (bottom) was generated by multiplying each temporal phosphorylation profile by the value obtained from the 5 minute mix quantification data for the particular stimulation condition. (D) Multiple tyrosine phosphorylation sites were identified and quantified from EGFR family member proteins. 4 representative phosphorylation sites are shown here, with response plots across temporal and conditional space. As can be seen by comparing HER2 pY877 and pY1248, different sites on a given protein are often differentially regulated.

obtained for 36 phosphorylation sites on 15 different proteins in the EGFR canonical signaling pathway (as defined by Yarden et al. (Yarden & Sliwkowski,

2001)) (Figure I.4). Coverage of the cell adhesion/migration pathway included quantitative information on 41 phosphorylation sites distributed along 16 proteins, including 9 tyrosine phosphorylation sites on δ -catenin.

One of the goals in the study presented in this chapter was to provide a more comprehensive definition of EGFR family signaling networks, including novel proteins or phosphorylation sites which may regulate differential cellular response to exogenous stimuli. To enable identification of novel phosphorylation sites, mass spectrometry analyses were performed in information dependent acquisition mode (automated selection of the most abundant species in a given full scan mass spectrum for MS/MS analysis), rather than targeting specific peptides and known phosphorylation sites for quantification. This data-dependent mode of operation was successful at identifying novel proteins and phosphorylation sites, as 122 of the 322 phosphorylation sites have not previously been described in the literature. Unfortunately, the use of automated ion selection to discover novel phosphorylation sites often precluded selection of low-abundance precursor ions for MS/MS fragmentation, and therefore temporal phosphorylation profiles were not obtained for all conditions for many peptides. In fact, 234 of the 322 sites in this work were detected and quantified in multiple analyses, but only 68 were quantified for all 4 stimulation conditions and the 5 minute comparison. This core group of 68 phosphopeptides, which include many of the key signaling nodes in the network, was then used for further computational analysis to obtain a mechanistic understanding of the effects of

HER-2 overexpression on cellular signaling networks and corresponding biological response to growth factor stimulation.

IV.3.2 Hierarchical Clustering reveals tyrosine phosphorylation sites co-regulated across conditional space

To identify groups of phosphorylation sites responding similarly across different cellular stimulations and to help define the role of novel phosphorylation sites within the EGFR signaling network, phosphorylation sites were clustered based on level of phosphorylation at a single time point, 5 minutes post-stimulation, for each cellular condition. Quantitative data for the 206 singly- or doubly-phosphorylated peptides obtained from replicate analyses of the 5 minute mix sample were normalized relative to EGF-stimulated P HMECs and hierarchical clustering was performed (using the Spotfire bioinformatics analysis program, see Methods), generating the heat map shown in Figure IV.2A.

Although many clusters of phosphorylation sites were generated by this analysis of the data, here we will discuss selected clusters that demonstrate the effectiveness of this approach in grouping co-regulated sites and thereby providing inference for functional categorization of novel sites. At the top of the map and highlighted in Figure IV.2B, a cluster of phosphorylation sites regulated primarily by EGFR activation was identified based on very low phosphorylation under HRG stimulation and high phosphorylation under EGF stimulation,

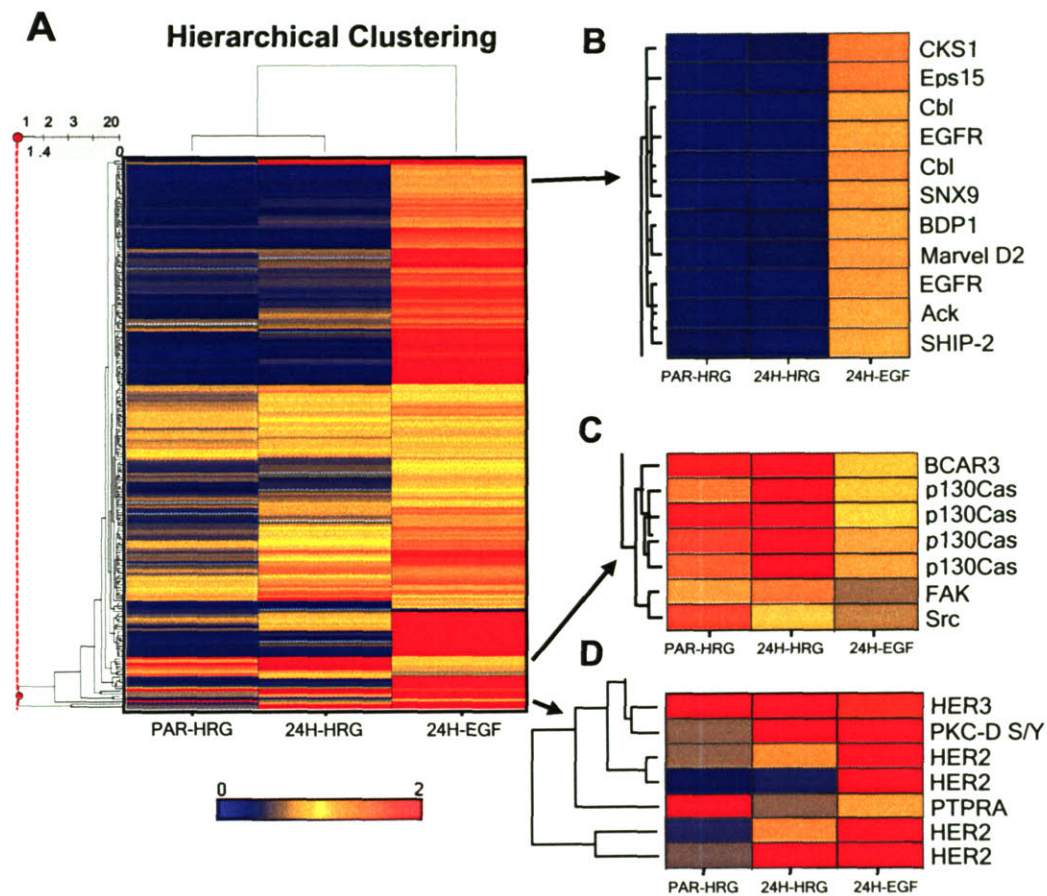


Figure IV.2 – Hierarchical Clustering of phosphorylated peptides to identify co-regulated phosphorylation sites. (A) Hierarchical clustering of the data generated a large heat map, in which both peptides (rows) and stimulation conditions (columns) were clustered. Note that the HRG-stimulated parental HMECs cluster with HRG-stimulated 24H cells, indicating that the level of phosphorylation for most of the sites depends more strongly on stimulation conditions (HRG vs. EGF) as compared to HER2 expression level. (B) Cluster of phosphorylated peptides downstream of EGFR activation, as indicated by high phosphorylation levels following EGF stimulation and low phosphorylation levels following HRG stimulation. (C) Cluster of phosphorylated peptides from proteins involved in the cell migration signaling network. Note the close association of multiple phosphorylation sites on p130Cas and BCAR3, two proteins from the same family. (D) Cluster of phosphorylated peptides primarily associated with high HER2 expression and including phosphorylation sites on HER2, HER3, and PKC- δ .

irrespective of HER2 expression level. This cluster includes phosphorylated peptides from proteins associated with the immediate-early response to EGFR stimulation and phosphorylated peptides from proteins involved in the internalization of EGFR, clusters which were also detected in the previous

chapter. Within this cluster the following phosphorylation sites were found: Y1068 and Y1148 of EGFR, which act as docking sites for GRB2 and SHC respectively (Batzer et al., 1994; Okabayashi et al., 1994), Y857/Y858 of Ack, a kinase previously reported to be phosphorylated upon EGFR stimulation (Sato et al., 1996; Taya et al., 2001), Y387 of BDP1, a protein phosphatase involved in the regulation of MAPK, Gab1 and HER2 (Gensler et al., 2004), and Y986 of SHIP-2, which has been shown to bind to SHC, inhibiting Rac activation and formation of [PtdIns(3,4,5)P3] and thus AKT activation downstream of EGFR stimulation (Coggeshall, 1998; Muraille et al., 1999; Pradhan & Coggeshall, 1997; Taylor et al., 2000). Also contained within this cluster are phosphorylation sites on two proteins whose role in EGFR signaling has not been characterized, Y7 of CKS1 and Y23 of Marvel D2. Although CKS1 has been shown to play a role in the downregulation of the cell-cycle regulatory protein p27Kip1 (Ganoth et al., 2001), and expression levels of CKS1 have been correlated with poor prognosis in breast cancer (Slotky et al., 2005), regulation of this protein in response to growth factor stimulation has not previously been characterized. Marvel D2 was related to EGFR signaling in the previous chapter. Although its role is not defined, it localizes toward the cell membrane (Sanchez-Pulido et al., 2002) and appears to be involved in the early response to EGFR activation. Finally, phosphorylation sites on 3 proteins related to internalization and ubiquitination of EGFR are also contained within this cluster, including Y177 of SNX9, which has been shown to be recruited for clathrin coated pit-mediated endocytosis and co-regulates EGFR degradation with Ack2 (Lin et al., 2002;

Soulet et al., 2005), Y849 of EPS15, which has been shown to be phosphorylated upon receptor activation by EGF and to induce receptor internalization (Salcini et al., 1999), and Y455 and Y552 of c-Cbl, an E3 ubiquitin ligase that binds EGFR Y1045 in the cytoplasm and endosomes, resulting in the targeting of EGFR for degradation (Grovdal et al., 2004).

Another cluster of co-regulated phosphorylation sites from the bottom of the heat map is highlighted in Figure IV.2C. This cluster contains protein phosphorylation sites which are phosphorylated preferentially by HRG stimulation of both parental and 24H cells and includes Y234, Y249, Y327 and Y387 in p130cas (breast cancer anti-estrogen resistance 1 (BCAR1)) and Y267 in BCAR3 (breast cancer anti-estrogen resistance 3). Interestingly, immediately contiguous to this cluster and with a similar activation profile are peptides containing putative Src phosphorylation sites, Y576 in FAK, and Y418 in Src, an autophosphorylation site (Roskoski, 2004). Both of these proteins have been shown to upregulate the migratory pathway (Yamboliev et al., 2001; Yang et al., 2005) and to interact directly with p130cas (Avraham et al., 2003; Riggins et al., 2003).

Three other small clusters highlighted in Figure IV.2D consist of phosphorylation sites on HER2, HER3, and two associated proteins. It is interesting to note that although HER2 sites may be clustered together, the distance between peptides in the clusters is large (scale has been compressed 3-fold due to space constraints), indicative of differential regulation at different sites and highlighting the need for site-specific quantification. In fact, HER2 site-

specific phosphorylation response to EGF or HRG stimulation may be informative as to the stimulation of specific signaling cascades downstream of activated heterodimers.

IV.3.3 Self-Organizing Maps define temporal and conditionally related clusters of phosphorylation sites

In order to identify clusters of tyrosine-phosphorylated peptides exhibiting similar temporal dynamics, as well as to globally visualize the high-dimensional information we have obtained, we used the self-organizing map (SOM) algorithm (Kohonen, 2001)¹. The SOM is a versatile clustering algorithm that transforms high-dimensional data into lower dimensional display, in a non-linear manner. Here we use the unified-distance matrix (or, U-matrix) approach, which allows robust identification of clusters, and the component plane representation which facilitates comparison of peptide phosphorylation response to exogenous stimuli (Kohonen, 2001; Ultsch & Siemon, 1989). Instead of taking a single map-unit as one cluster, we use U-matrix to identify groups of map-units that together comprehend a cluster. The U-matrix illustrates the mean distances between neighboring map units after the SOM training phase. These distances are color-coded so that close proximity of two map-units is colored with bluish colors, while

¹ The SOM algorithm used in this chapter is available online as a MATLAB™ toolbox (<http://www.cis.hut.fi/projects/somtoolbox/>). We choose to use this algorithm in stead of the Spotfire™ algorithm used in chapter I, because of its more rigorous mathematical construction.

shades of yellow and red denote dissimilarity. Clusters can be identified as continuous bluish regions (valleys) surrounded by yellow or red “mountains.”

We applied this SOM algorithm to 62 “core” phosphorylated peptides for which temporal profiles were generated under all conditions (six of the “core” phosphorylated peptides had minimal response across all conditions and time points and were removed from this analysis). Data were mean centered and unit variance scaled (z-score normalized) across all time-points and conditions. From the U-matrix in Figure IV.3 four clusters are readily identifiable. Statistical significances for these clusters were computed with a permutation test based method (Hautaniemi et al., 2003) and the clusters were found to be statistically significant ($p < 0.05$). The entire SOM display including the component planes is shown in Figure IV.4.

By comparing the individual phosphorylation profiles (blue dashed lines) or average phosphorylation profile (red line) within each cluster (Figure IV.3), it is clear that the algorithm has clustered peptides whose phosphorylation level is increased under selected conditions, information which can be used to link phosphorylation sites within a cluster to activated receptor homo- or heterodimers. For instance, the first cluster (c1 in Figure IV.3) consists of 18 peptides whose phosphorylation levels, on average, are highest following EGF stimulation of P or 24H HMECs, conditions which would lead to activation of EGFR homodimers or EGFR-HER2 heterodimers, respectively. Correspondingly, most of the proteins in this cluster have been previously associated with proliferation and early response to EGF stimulation, including EGFR Y1068 and

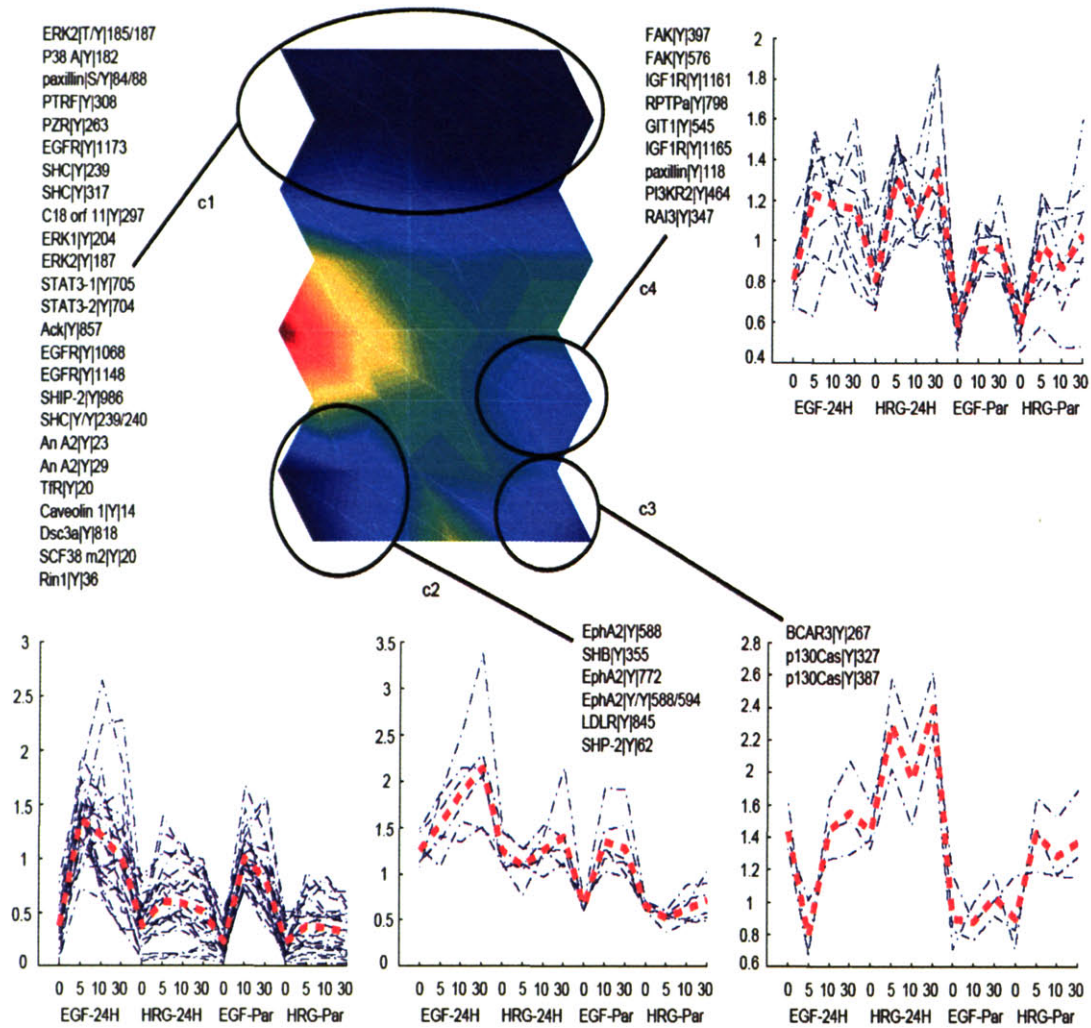


Figure IV.3 – Self Organizing Map to cluster phosphorylation sites across conditional and temporal space. Phosphorylation levels for 62 core phosphorylated peptides, which were detected and quantified at 4 time points for each of the 4 stimulation conditions and the 5 minute mix analysis, were normalized and placed into a self-organizing map (SOM). Four statistically significant clusters were identified in the SOM (circled bluish valleys surrounded by yellow “mountains”). The components for each of the four clusters are indicated in the Figure along with the phosphorylation profiles (blue dashed lines) and an average profile for each cluster (red line) across conditions and time-points (16 dimensions).

Y1148, STAT-3 Y705 (STAT-3 isoform 1) and Y704 (STAT-3 isoform 2), SHIP-2 Y986, SHC Y239, Y240 and Y317 and early effectors downstream of them, including MAP kinases ERK2 (phosphorylated at Y187 and at T185 and Y187), Erk1 Y204, and p38 α Y182.

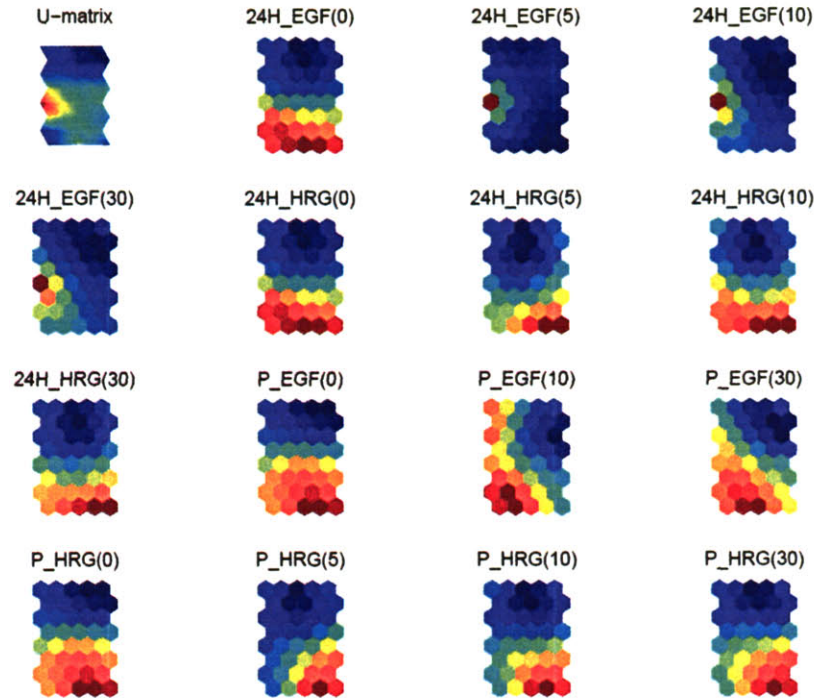


Figure IV.4 – Self Organizing Map (SOM) Component planes. Phosphorylation levels for 62 core phosphorylated peptides, which were detected and quantified at 4 time points for each of the 4 stimulation conditions and the 5 minute mix analysis, were normalized and placed into a self-organizing map (SOM). The component planes show the explicit dynamic phosphorylation pattern of the peptides clustered in each neuron across the complete battery of biological conditions (normalized to P_EGF(5)). Blue neurons represent low phosphorylation for the peptides at that time and condition (with respect to P_EGF(5)) and red neurons represent high phosphorylation for the peptides at that time and condition (with respect to P_EGF(5)).

In comparison to c1, the three other clusters in the U-matrix (c2-c4) contain peptides whose phosphorylation appears to be regulated by HER2 activation, as phosphorylation levels on these peptides are highest when HER2 is overexpressed. More specifically, peptides in the second cluster (c2) are primarily phosphorylated downstream of EGFR-HER2 heterodimers, since the highest levels of phosphorylation occur in the 24H HMECs stimulated with EGF. This cluster consists of SHB Y355, SHP-2 Y62, LDL receptor Y845, and three

EphA2 tyrosine sites, including the activation loop at Tyr 772 (Kinch & Carles-Kinch, 2003). Of these proteins, only SHP-2 has been previously associated with EGFR activation (Qu et al., 1999). Interestingly, each of the other proteins has been associated with the cell migration response to VEGF stimulation (Cheng et al., 2002; Holmqvist et al., 2003; Prager et al., 2004), a response which requires EGF autocrine signaling following VEGF stimulation in endothelial cells (Semino et al., 2006). Our data links these phosphorylation sites to the EGFR family signaling network, and as demonstrated below, shows strong correlation between these phosphorylation sites and cell migration in response to EGF stimulation of 24H HMECs.

The third cluster (c3), contains 3 peptides that are primarily phosphorylated downstream of activated HER2-HER3 heterodimers, as implicated by maximal phosphorylation levels following HRG stimulation of 24H HMECs. This cluster includes phosphorylation of p130Cas at Y327 and Y387, and phosphorylation of BCAR3 at Y267. Both of these proteins are part of the same family and have been shown to regulate or be regulated by Src activity in the cell migration signaling network. Similar to c2 and c3, peptides in the fourth cluster (c4) are predominantly phosphorylated downstream of HER2, but are equally phosphorylated downstream of EGFR-HER2 or HER2-HER3 heterodimers or possibly activated through active HER2 homodimers (similar phosphorylation levels are seen in EGF or HRG stimulated 24H HMECs). This cluster includes phosphorylation of retinoic acid induced protein 3 (RAI3) Y347, paxillin Y118, GIT1 Y545, FAK Y397 and Y576, receptor protein tyrosine

phosphatase alpha (RPTP- α) Y798, PI3K Y464, and Insulin-like growth factor receptor (IGF-1R) Y1161 and Y1165. A complex containing GIT1 and paxillin has been shown to regulate cell motility, possibly through recruitment of PAK or PIX to the leading edge (Manabe et al., 2002), although the role of tyrosine phosphorylation in this process has not been established. Other proteins in this cluster have also been shown to be co-regulated, as RPTP- α has recently been shown to regulate integrin signaling through Src activation in a PI3K-dependent manner downstream of insulin activation (Vulin et al., 2005), and Src activation is known to regulate tyrosine phosphorylation of FAK. It is worth noting that the phosphorylation sites on IGF-1R (Y1161 and Y1165) occur within a peptide that is homologous with the Insulin receptor. Although not conclusive, assignment to IGF-1R is based on the presence of a phosphorylation site on a peptide specific for IGF-1R in the larger data set and greater levels of IGF-1R as compared to Insulin receptor on HMEC parental and 24H cells (H.S. Wiley, personal communication). Activation of IGF-1R has been found to be correlated with increased motility in a breast cancer cell line (Zhang et al., 2005b); in our system, increased phosphorylation is most likely associated with autocrine release and binding of IGF-1 following stimulation of EGFR family members.

These clusters make clear that analyzing phosphorylation sites across 16 dimensions of temporal and conditional space can reveal co-regulated phosphorylation sites. Moreover this information can be used to connect phosphorylation of specific sites to activation of EGFR family member dimers. Although both clusters 3 and 4 contain phosphorylation sites from proteins known

to regulate migration, they are statistically distinct clusters whose differential response to exogenous stimuli may be due to protein sub-cellular localization, as proteins from cluster 4 are known to be associated with the membrane, while proteins in cluster 3 are primarily cytosolic. SOM-based data analysis can reveal interesting hypotheses and connectivity in the data, such as the potential role of GIT1 Y545 affecting cell migration or the inclusion of PTPR- α , PI3K, and IGF-1R phosphorylation sites in a single cluster, but it is still necessary to relate phosphorylation and phenotypic data to solidify these hypotheses; thus we proceed to measure proliferation and migration in our system.

IV.3.4 Cell proliferation and migration are differentially stimulated via EGFR and HER2

In order to correlate signaling data with cellular response, we quantified both cell migration and cell proliferation in the HMEC parental and 24H cell lines. Measurements were obtained under serum-free, HRG (80 ng/ml), or EGF (100 ng/ml) stimulating conditions.

Cell migration was measured using a wound-closure assay adapted for a 96-well plate format and automated fluorescent imaging. Under all conditions investigated, the 24H cells moved more rapidly into the induced wound, thereby reducing the original wound area at a greater rate than the parental cells (Figure IV.5 A-C). The wound closure trajectories were fit to a line, and the slopes were

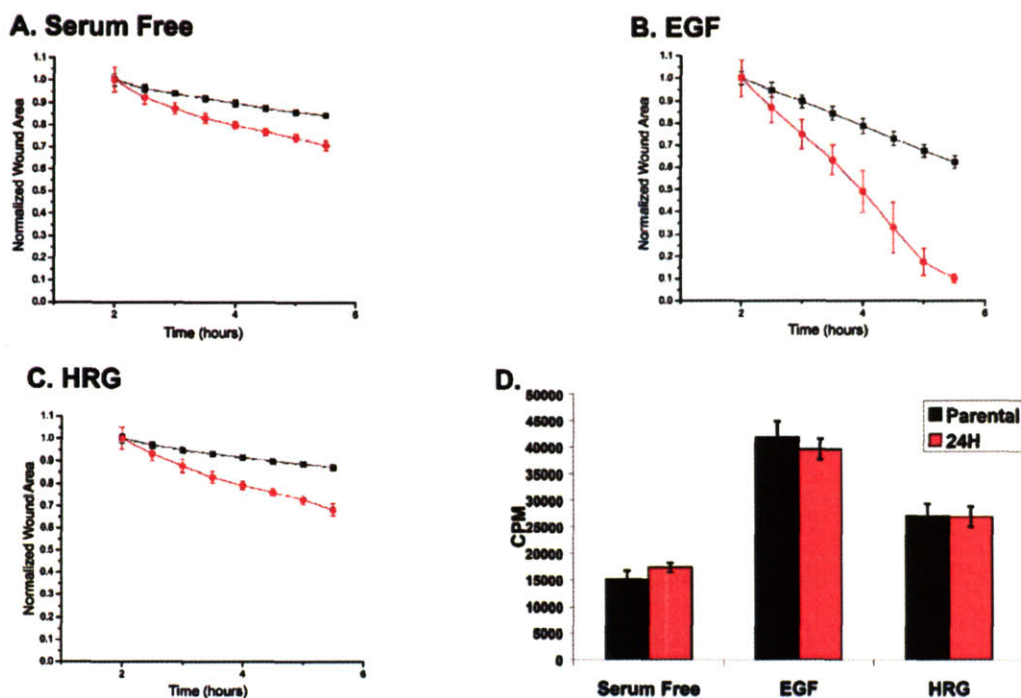


Figure IV.5 – EGF and HRG drive migration and proliferation to varying extents in HMEC parental and HER2 over-expressing cells. Parental (black) and 24H (red) wound healing data shown for: (A) serum-free conditions, (B) treatment with 100 ng/ml EGF, and (C) treatment with 80 ng/ml HRG. (D) Proliferation, as measured by $[^3\text{H}]$ thymidine incorporation, for both cell types under the treatment conditions shown in (A)-(C).

the input for partial least squares modeling. Interestingly, the greatest difference between the two cell lines occurs during EGF treatment. By comparison, HRG stimulation, while inducing a slightly higher rate of wound closure for 24H HMEC relative to P HMEC, does not seem to drive significantly enhanced migration relative to serum-free conditions.

Cell proliferation was measured by $[^3\text{H}]$ thymidine uptake 25 hours after ligand stimulation. Figure IV.5 D shows that EGF treatment increased thymidine uptake to a greater extent than did HRG treatment, but both treatments induced higher amounts of proliferation than seen in serum-free conditions. In direct

contrast to the migration phenotype, there was no significant difference (i.e., $p \gg 0.05$ in all cases) between the two cell types measured under any of the stimulation conditions. Thus, HER2 overexpression did not seem to facilitate enhanced proliferation in any of the conditions probed.

The fact that HER2 over-expression mediated differences in migration, particularly under EGF stimulating conditions, but did not do so for proliferation indicates that some of the signaling molecules differentially regulated by HER2 expression levels play a role in driving higher levels of migration while at the same time remaining agnostic to cell proliferation. The linear regression modeling discussed below integrates our quantitative migration, proliferation, and signaling data to describe, among other things, a set of signaling molecules that is most relevant for the changes in migration induced by HER2 overexpression.

IV.3.5 Modulation of EGFR signaling network by HER2

To assess the effect of increased HER2 expression levels in the canonical EGF activated pathway, the phosphorylation level for sites observed in EGF-stimulated 24H cells was divided by the phosphorylation level for the same site and stimulation time in EGF-stimulated parental cells, producing a fold change in phosphorylation level for a given site and time. A subset of the proteins and phosphorylation sites within the canonical EGFR signaling network are shown in Figure IV.6. As is evident from this figure, increased HER2 expression affects

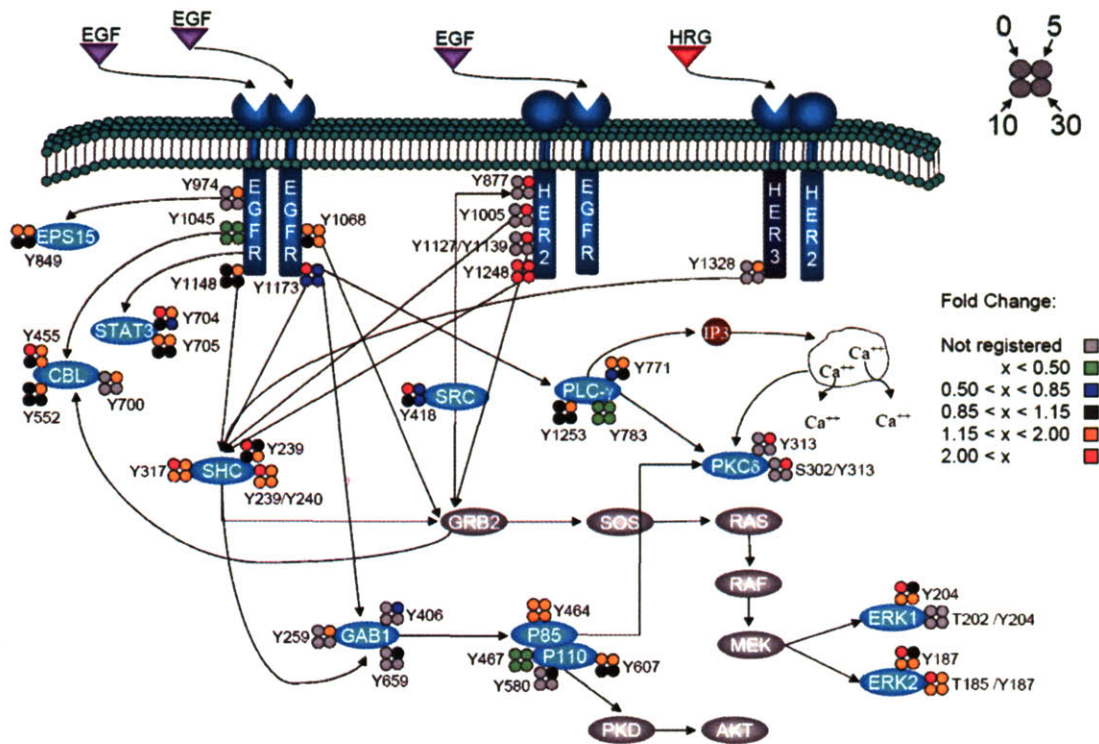


Figure IV.6 – Effect of increased HER2 expression on phosphorylation sites within the EGFR signaling network. Visualization of the fold change in phosphorylation level between 24H and parental cells stimulated with EGF provides a network view of the mechanistic effects of HER2 overexpression on the canonical EGFR signaling cascades.

most phosphorylation sites on selected proteins in the EGFR signaling network, but not all phosphorylation sites on a given protein react equally to this perturbation. For instance, each of the multiple phosphorylation sites on EGFR exhibit different regulation at low or high HER2 expression levels, including increased phosphorylation of Y974 and decreased phosphorylation on Y1045 under HER2 overexpression as compared with basal HER2 expression. Both of these sites appear to regulate receptor internalization and degradation: Y974A or Y974F mutations have been shown to decrease receptor internalization rates (Sorkin et al., 1996), and Cbl (E3-ubiquitin ligase) binding to EGFR Y1045 is

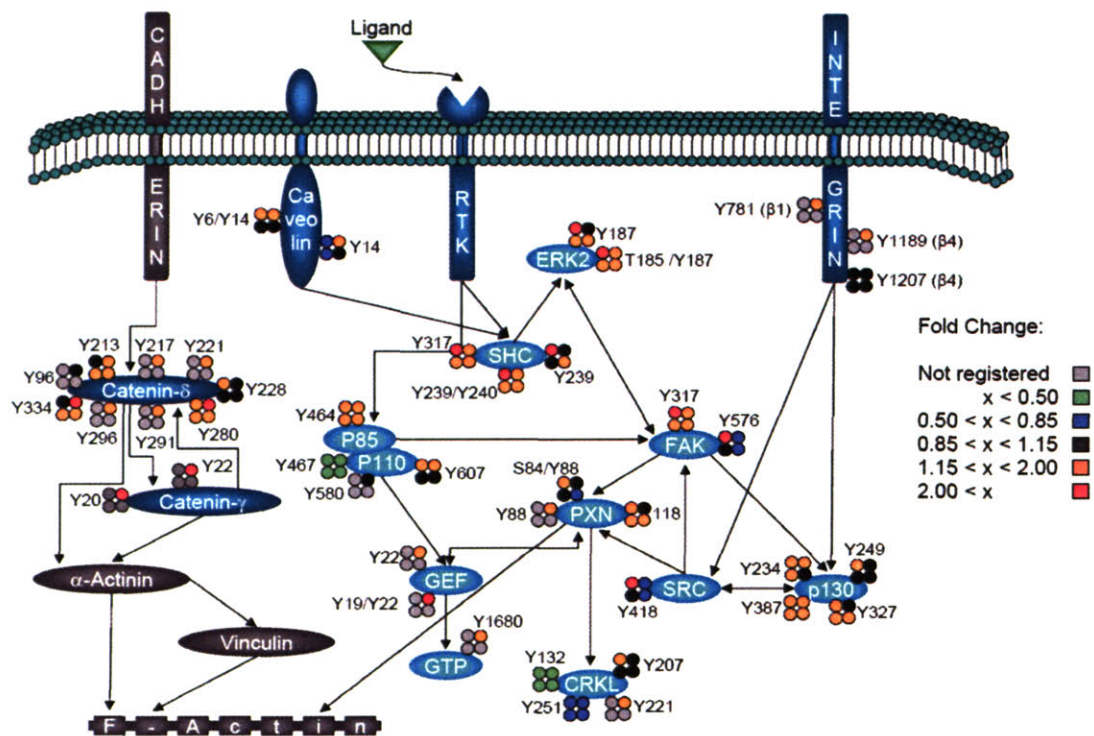


Figure IV.7 – Effect of increased HER2 expression on phosphorylation sites within the EGFR signaling network 2. Visualization of the fold change in phosphorylation level between 24H and parental cells stimulated with EGF provides a network view of the mechanistic effects of HER2 overexpression on the cell migration associated signaling pathway.

required for lysosomal sorting and receptor degradation (Grovdal et al., 2004). Decreased phosphorylation at the Y1045 site should lead to decreased ubiquitination of activated EGFR, thereby providing a mechanism for observed increase in recycling of activated EGFR in the 24H cell line relative to parental 184A1 HMECs (Hendriks et al., 2003b).

It is also interesting to note the higher basal phosphorylation level for selected sites in the 24H cell lines relative to parental HMECs. Increased basal phosphorylation associated with HER2 overexpression could be mediated by HER2 autophosphorylation, cross-phosphorylation of EGFR in the absence of

ligands, or even through increased autocrine stimulation of HER2 heterodimers. For most sites with high levels of basal phosphorylation in the 24H cells, stimulation with saturating concentrations of EGF typically resulted in a greater response in parental cells relative to 24H cells, such that after EGF stimulation for 5 minutes many of the sites had similar levels of phosphorylation in both cell lines. A specific example of this behavior is provided by phosphorylation of the ERK2 activation loop (Y185 and T185/Y187). Much higher levels of basal phosphorylation were detected in the 24H cells relative to parental cells, but only a slight difference in phosphorylation at these sites remained after 5 minutes of EGF stimulation. These results are consistent with our proliferation data, in which no significant difference between 24H and parental cells under EGF stimulation was observed, in contrast to serum-free 24H cells, which had greater proliferation than parental HMECs ($p < 0.1$). These results are also consistent with previous literature demonstrating ERK2's role as a potent activator of proliferation (Cobb et al., 1994). In fact, many protein phosphorylation sites associated with proliferation behave similarly to ERK2 in our system, including STAT-3 Y705 (and Y704 from STAT-3 isoform 2), EGFR Y1173, and SHC Y239, Y240, and Y317.

Since HER2 over-expression has been found to affect cell motility in our study here as well as in previous work (Spencer et al., 2000), the effect on protein phosphorylation in a subset of the pathways related to cell migration is shown in Figure IV.7. From this figure, it is clear that there is an overall increase in phosphorylation of many of these pathway proteins in the presence of HER2

overexpression. Perhaps most striking is the increase in phosphorylation for all of the phosphorylation sites detected on catenin- δ and catenin- γ . Catenins are known to interact with E-Cadherin, the main cell-cell adhesion protein in epithelial cells (Davis et al., 2003). Catenin- δ , a member of the p120 catenin family, has been shown to regulate E-Cadherin turnover by modulating its internalization and degradation rate (Davis et al., 2003), stabilizing it on the cell surface when bound to it. HER2 overexpression in breast carcinomas inhibits the transcription of E-Cadherin (D'Souza & Taylor-Papadimitriou, 1994) and has also been found to destabilize the catenin-cadherin complex, leading to decreased adhesion (Jawhari et al., 1999). Also, EGFR has been found to directly phosphorylate p120 catenin at Y228 (Mariner et al., 2004), and EGFR inhibition was found to promote assembly of cell adhesions (Lorch et al., 2004). The accumulated evidence of these reports is consistent with our data as it argues that Src- or RTK-related phosphorylation of catenin- δ leads to separation of the catenin-cadherin complex, resulting in destabilization of E-Cadherin and an increase in cell migration. Tyrosine phosphorylation of catenin- γ has also been related to loss of cell-cell adhesions in EGF-stimulated, E-Cadherin positive, cervical cancer cells (Moon et al., 2001). Our data shows high levels of phosphorylation of catenin- δ and γ in HER2 over-expressing cells under EGF stimulation, directly contributing to loss of cell adhesion and an increase in cell motility.

Although catenin phosphorylation and loss of E-cadherin-based cell-cell adhesion may provide part of the migratory response following EGF stimulation in 24H cells, increased phosphorylation of many additional components of the

cell migration pathways are also likely to be contributing to increased migration of these cells relative to parental cells. For instance, FAK, Src, Paxillin and p130Cas have all been shown to interact with each other, to localize at the focal adhesions, and to play a fundamental role in the actin cytoskeleton reorganization and motility pathways (Webb et al., 2004). Phosphorylation of paxillin (Y31) and (Y118) is known to regulate membrane spreading and ruffling in cell migration and adhesion (Tsubouchi et al., 2002). FAK and Src are two of the major kinases involved in cell migration. FAK (Y397) is the major autophosphorylation site and acts as a docking site for Src and PI3K (Cary & Guan, 1999); this site is also necessary for both p130cas and paxillin phosphorylation in response to FAK expression (Cary & Guan, 1999). Src Y418 is a major Src autophosphorylation site, whose phosphorylation results in self-activation (Roskoski, 2004). Src can mediate FAK related migration by further phosphorylation of paxillin or p130Cas, which may result in focal adhesion turnover and/or lamellapodia and filipodia formation via either the RAC, JNK pathway or the RhoGAP pathway (Playford & Schaller, 2004). We conclude that the effect of increased HER-2 expression levels is partly increased interaction of FAK, Src, P130Cas and paxillin, a small but noticeable increase in ERK2 activity and the upregulation of catenin phosphorylation and thus E-cadherin turnover, all combining to drive an increased migratory response in the 24H cell line (as in Figure IV.5).

IV.3.6 HRG vs EGF stimulation in the presence of HER2

In addition to investigating the effects of HER2 on these pathways in Figures IV.6 and IV.7, our data also enabled comparison of signaling downstream of activated HER2:HER3 vs. EGFR:HER2 heterodimers, resulting from stimulation of the 24H cells with HRG and EGF, respectively. For this analysis, the phosphorylation level resulting from HRG stimulation of 24H cells was divided by the phosphorylation level resulting from EGF stimulation of 24H cells to produce the fold change ratio for each phosphorylation site in the canonical signaling pathways (Figure IV.8). As expected, stimulation with HRG instead of EGF caused a significant increase in HER3 phosphorylation and a large decrease in EGFR phosphorylation. Perhaps not surprisingly given the relative receptor expression levels (20,000-30,000 copies/cell for HER3 vs. 200,000 copies/cell for EGFR) and the kinase-dead nature of HER3, stimulation with saturating concentrations of HRG resulted in decreased phosphorylation of almost all downstream proteins as compared to stimulation with saturating concentrations of EGF (Figure IV.8). Specifically, most of the downstream effectors which lead to proliferation (STAT-3 Y705 (Y704), ERK T202/Y204, all three phosphorylation sites on SHC) were phosphorylated to a lesser degree under HRG stimulation, correlating with decreased proliferation in HRG stimulated 24H cells as compared to EGF stimulated 24H cells (as in Figure IV.5). Other sites primarily associated with migration (Src, PKC- δ , and PI3K) do

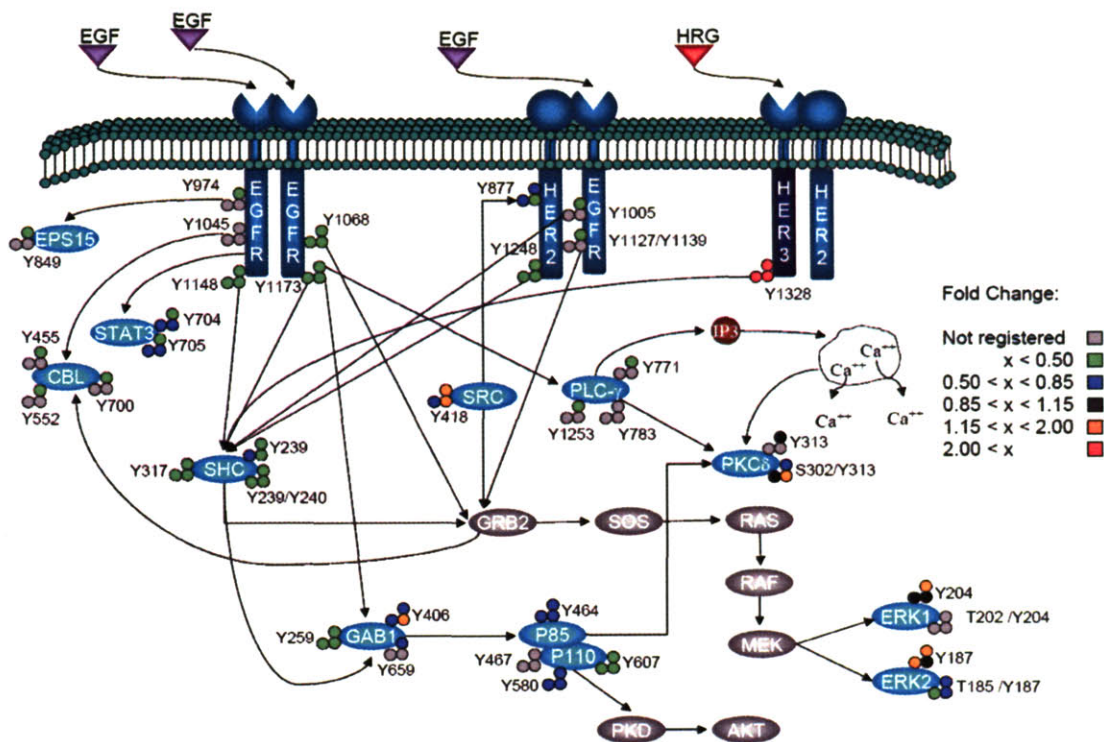


Figure IV.8 – Effect of HRG stimulation versus EGF stimulation in high HER2 expressing cells. Visualization of the fold change in phosphorylation level between HRG-stimulated and EGF-stimulated 24H cells provides a network view of the mechanistic effects underlying differential response to distinct growth factor stimuli on the canonical EGFR signaling cascades.

not show the same decrease in phosphorylation when comparing HRG to EGF stimulation of these cells. In fact, both Src and PKC- δ show an increase in phosphorylation under HRG activation at 5 and 10 minutes respectively which coupled with our response data indicates that signaling distinct from Src, PKC- δ , and PI3K controlled pathways may govern cell migration in our system.

To analyze the effect of HRG stimulation on the cell migration signaling network, signaling downstream of activated HER2:HER3 vs. EGFR:HER2 heterodimers was also compared for selected proteins within a subset of the pathways related to cell motility (Figure IV.9). Comparing phosphorylation levels

for specific sites under different stimulation conditions provides a potential hypothesis to explain the mechanism underlying the quantitative phenotypic migration measurements, in which EGF stimulation resulted in a dramatic increase in migration relative to HRG stimulation of 24H cells (see Figure IV.5). Given the difference in migration rates between these two cell states, it is not surprising to note that protein phosphorylation sites on almost all effectors of migration have decreased phosphorylation levels following HRG stimulation as compared to EGF stimulation. However, in contrast to most proteins in the motility network, tyrosine phosphorylation levels on FAK, p130Cas, Src and paxillin were increased following HRG stimulation compared to EGF stimulation, indicating that the slight migratory effect of HRG stimulation of 24H cells is directed primarily through amplified phosphorylation of a very specific pathway driven through these four proteins.

Taken together, the data from Figures IV.6, IV.7, IV.8 and IV.9 demonstrate that 24H cell migration can be induced by either broad upregulation of the migratory pathway, including loss of cell adhesions (24H cells under EGF stimulation) or, to a much lesser extent, by more intense upregulation of a restricted subset of the migratory pathway (as in 24H cells stimulated with HRG). By comparison, 24H cell proliferation was induced to a greater extent by EGF stimulation as determined by phenotypic assay; correspondingly, phosphorylation levels were higher on almost all proliferation-associated proteins on 24H cells stimulated with EGF as compared to 24H cells stimulated with HRG. Also consistent with the phenotypic data, the effect of increasing HER2 expression

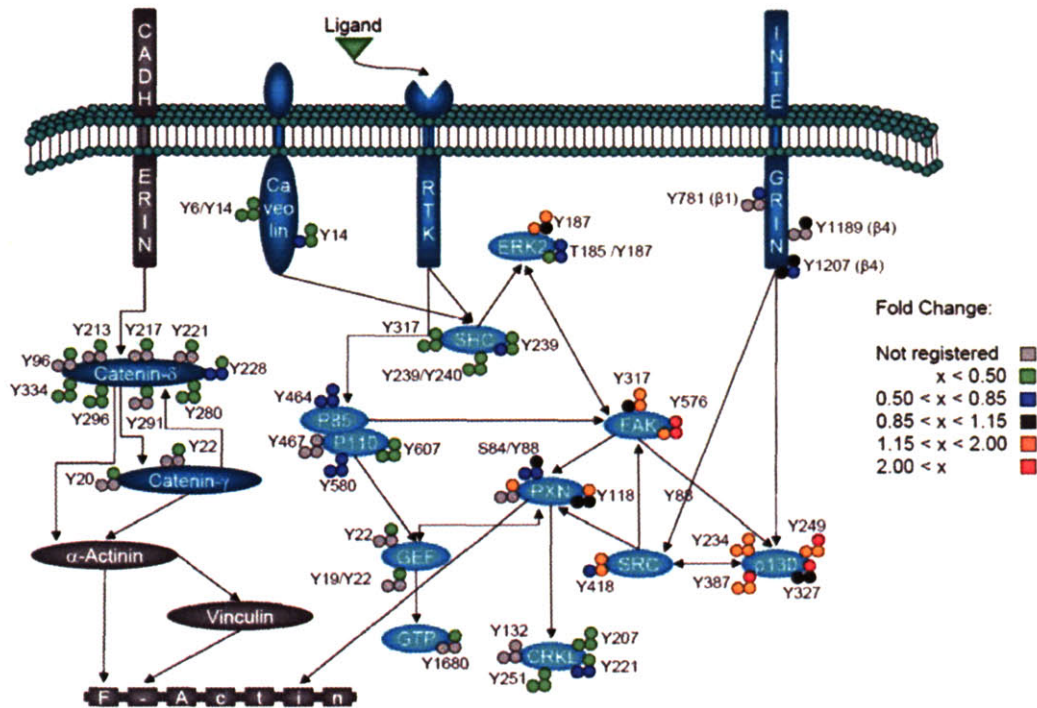


Figure IV.9 – Effect of HRG stimulation versus EGF stimulation in high HER2 expressing cells 2. Visualization of the fold change in phosphorylation level between HRG-stimulated and EGF-stimulated 24H cells provides a network view of the mechanistic effects underlying differential response to distinct growth factor stimuli on the cell migration associated signaling pathway.

levels was most dramatic for proteins in the cell migration signaling network, as most proteins displayed a sustained increase in phosphorylation levels, but was much less dramatic for proteins in the proliferation-associated network, as increased basal phosphorylation levels in 24H cells were not sustained following EGF stimulation.

IV.3.7 PLSR modeling correlates signals with cell functional responses

We have constructed a model using partial least-squares regression (PLSR), a method we have previously found to be effective in relating cell signaling data to cell behavioral response data in a quantitative and integrative manner (Janes et al., 2005). Information obtained through our proteomic studies was represented in an $M \times N$ matrix (the X-block), where M is the number of conditions investigated, and N is the number of peptide metrics measured. An entry in the matrix with coordinates (i,j) represents the column j metric (i.e. ERK Y187 phosphorylation at 5 minutes) measured under the row i condition (i.e. parental cell line treated with EGF). For each condition, the metrics included in the model were phosphorylation measurements at 5, 10, and 30 minutes in addition to the integral of this time course (with integrals being used as a measurement for the 'net' phosphorylation over the 30 minute time course). Cell behavior measurements comprised an $M \times P$ matrix (the Y-block), where M is again the number of conditions and P is the number of behavior measurements obtained. Partial least-squares regression analysis produced a vector of coefficients indicating the importance of each signaling metric with respect to cellular behavior. In addition, PLSR provided a reduced-dimension map, with axes defined as linear combinations of our original signaling metrics (Figure IV.10A), on which both signals and cellular behavior can be represented. Figure IV.10A shows that our original dataset, consisting of 248 dimensions (i.e., 248 phosphorylation signal metrics), has been reduced to 3 dimensions using PLSR,

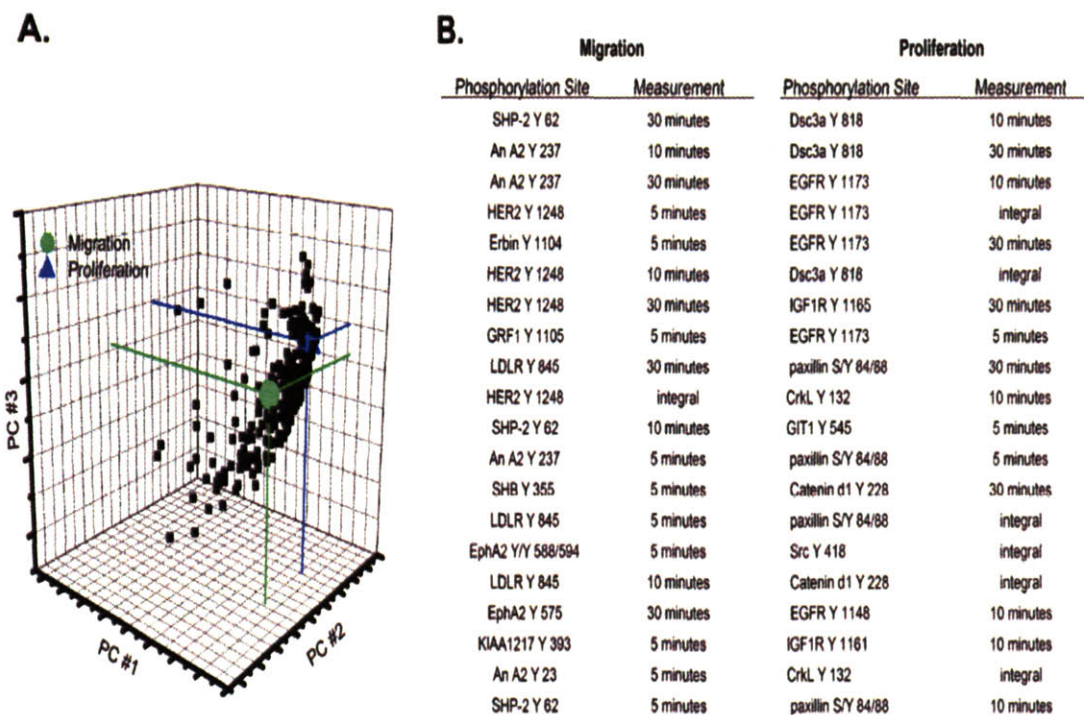


Figure IV.10 – Partial least squares regression correlates 248 protein metrics to cell migration and proliferation. (A) Visual representation of a reduced (3) dimension graph showing all 248 protein measurements and the cellular outputs with cell migration (green), cell proliferation (blue), and protein signals (black). **(B)** List of the top 20 signal-behavior covariates for both migration and proliferation.

each of which incorporates a quantitative combination of multiple signals. The projection of an individual signal in the direction of a given cellular behavior in the PLS space determines how important the phosphorylation signal is to the behavior. In Figure IV.10B, we list the top 20 signals that positively correlate most strongly with each cell behavior. Importantly, even though we can identify small sets of variables that correlate strongly with each cellular output, 148 out of the 248 protein metrics had a variable importance for projection (VIP) value of greater than 1, indicating that these 148 protein metrics play an important role in our model (see Materials and Methods for VIP calculation). This highlights the

great advantage of proteome measurements that quantitatively capture dynamic information flow through a large number of nodes. Our model was validated through cross-validation and had a goodness of prediction (Q^2) of 0.89 (see Materials and Methods). From the top 20 signal behavior co-variates for proliferation and migration in our model (Figure IV.10B), HER2 Y1248 appears to be the main upstream activator of migration while proliferation seems to be activated by EGFR Y1173 and to a lesser extent EGFR Y1148. This finding correlates with literature showing that the presence of HER2 Y1248 is necessary to induce migration of breast cancer cell lines (Dittmar et al., 2002) whereas both EGFR Y1148 and Y1173 are Shc binding sites and thus are able to activate the ERK signaling pathway, thereby driving proliferation (Okabayashi et al., 1994). In this model, SHP-2 Y62 and Annexin A2 Y237, among others, also appear to be highly correlated with migration. Both of these proteins have previously been implicated in regulating migration, as SHP-2 has been shown to promote migration in MCF-7 cells by inducing loss of E-Cadherin expression and production of matrix metalloproteinase 9 (Wang et al., 2005a), while Annexin A2 has been shown to mediate mitogenesis, cell migration and loss of focal adhesions when activated by tenascin-c (Chung et al., 1996; Liu et al., 2003) and it has also been identified as a major substrate for EGFR phosphorylation (Rothhut, 1997). Interestingly, treating cells with withaferin, a novel inhibitor of Annexin A2, has recently been shown to decrease migration, further validating the correlation between Annexin A2 and cell migration (Falsey et al., 2006).

In addition to EGFR Y1173 and Y1148, desmocollin (Dsc3a) Y818, IGF-1R Y1165, and catenin- δ Y228 (among others) are most strongly correlated with proliferation. IGF-1R is known to induce cell proliferation and survival through activation of the MAPK and PI3K pathways (Wang et al., 2005b), and has also been associated with EGFR signaling in Tamoxifen resistant breast cancer cell lines, potentiating their mitogenic strength (Knowlden et al., 2003). The correlation of Dsc3a and catenin- δ to cellular proliferation is initially surprising, as both of these proteins are associated with cell adhesions, leading to the expectation that they should correlate to migration instead of proliferation. However, EGFR inhibition has been shown to increase the presence of desmocollin and desmoglein at cell-cell borders (Lorch et al., 2004), and EGFR has also been shown to signal to E-cadherin complexes through phosphorylation of catenin- δ Y228 (Mariner et al., 2004). Since the conditional and temporal profiles for both of these sites (Dsc3a Y818 and catenin- δ Y228) display strong similarity to that observed for EGFR Y1173, it can be argued that they are either directly or closely downstream of EGFR and that phosphorylation of these sites may destabilize cell-cell adhesions, a step needed for both proliferation and migration. Further biological experiments are clearly needed to more firmly establish the functional roles of these proteins in regulating cellular proliferation under these stimulation conditions.

It is worth noting that a large proportion of the phosphorylation events contained in our regression model influence both migration and proliferation to some extent. Importantly, the model provides a quantitative metric describing the

strength of correlation for each phosphorylation site relative to migration or proliferation; data which can be used to design further experiments to specifically perturb selected biological outcomes -- i.e., to inhibit migration one might want to target HER2, SHP-2 and EphA2. This type of data-driven modeling can also provide insight to the functionality of unknown proteins such as KIAA127; in our model, phosphorylation of this protein at Y393 correlates strongly to migration. Interestingly, KIAA1217 is a novel protein highly homologous to p140CAP (p130Cas-associated protein), which has been shown to be tyrosine phosphorylated in response to EGF stimulation and to participate in actin cytoskeleton organization and cell spreading (Mariner et al., 2004).

Combining large-scale quantitative analysis of tyrosine phosphorylation with quantitative phenotypic measurements has provided the means with which to understand the relationship between these phosphorylation events and their relation to cellular responses. The findings we have discovered here illustrate how HER2 over-expression influences signaling network activities important for governing cell proliferation and migration behavior, common to and distinct between EGFR-binding ligand and HER3-binding ligand conditions. This quantitative information offers unusual opportunity for understanding prospective drug effects in a network-wide manner, and may offer novel targets for intervening in biological processes downstream of activated receptor tyrosine kinases.

IV.4 DISCUSSION

Through coupling time-resolved proteomic analysis of tyrosine phosphorylation across four cellular conditions of HER2 expression level and EGF family ligand treatment, with phenotypic profiling of the cells according to migration and proliferation levels under the same conditions and applying several different computational analyses to the experimental results we have quantified the effect of increased HER2 expression on cellular signaling networks and cellular response downstream of EGF or HRG stimulation. Furthermore, we have been able to find strong correlative behavior between the signals and cellular phenotype.

By performing the phosphorylation analysis in information dependant acquisition mode (discovery mode), we were able to identify and quantify a large number of well-characterized and novel phosphorylation sites in the EGFR family signaling network. However, the price for high network coverage was paid by the low reproducibility of phosphorylation sites across conditions. Of a total of 332 phosphorylated peptides quantified only 68 (20.8%) were found under all biological conditions. From this subset we took a core of 62 peptides that show conditional variation for subsequent computational analysis. Notwithstanding the low reproducibility of peptides across conditional space the core dataset still remain to our knowledge the most complete overview on the effects of HER2 on signal transduction networks.

In order to ascribe potential functionality to novel sites in the data set, we applied both hierarchical clustering and self-organizing maps to identify clusters

of phosphorylation sites co-regulated across conditional (hierarchical clustering of all sites found in the 5 minute runs) or temporal and conditional space (SOM analysis of the 62 phosphorylation sites in the core dataset). Although these clusters reveal dynamic co-regulated modules within the network, and are helpful to track which set of EGFR family homo- or heterodimers are driving different subsets of the observed cellular signaling networks they were not enough to understand how HER2 expression was driving biological variation in the cells.

To gain insight into the specific effects of HER2 expression in cellular responses, we generated figures to visualize conditional and temporal changes in protein phosphorylation levels for proteins associated with selected signaling pathways. This level of analysis revealed several interesting hypotheses, such as the potential role of decreased phosphorylation on EGFR Y1045 leading to reduced receptor endocytosis and recycling rates with increased HER2 expression, the involvement of FAK, Src, p130Cas and paxillin driving migration downstream of activated HER2-HER3 heterodimers (following HRG stimulation in 24H cells) or the importance of the catenin network in cell migration downstream of activated EGFR-HER2 heterodimers (following EGF stimulation of 24H cells). However interesting, the insights gained were intuitive and lacked rigorous quantification allowing more definite conclusions on the role of different phosphorylation sites in the measured cellular phenotypes.

Finally, to quantitatively define the relationship between our core phosphoproteomic dataset and proliferation and migration we used partial least square regression analysis. In fact, we were able to construct a predictive model

that allowed us to find the phosphorylation events correlating the closest to both migration and proliferation. This approach resulted in the correlation of novel and expected sites to migration and proliferation.

Overall, the work presented in this chapter accomplished several goals. We generated what it is to date the largest dynamic and conditional dataset of signal transduction downstream of EGFR family members activation. Moreover, by using computational analysis we were able to discriminate the specific dimers upstream of many of the measured signals and we were also able to correlate those signals to two cellular outcomes (migration and proliferation). As a continuation of this project Neil Kumar has deepened his work in the computational analysis of the core dataset and its relationship to cellular phenotype with the objective of quantifying the difference between the cellular systems and finding the most predictive signals in the dataset for cell behavior. This analysis has been submitted for publication.

A new project emerging from the results of this chapter deals with the problem of mass spectrometry run reproducibility. In order to obtain large, quantitative, statistically significant datasets for further mathematical analysis it is necessary to have at least several measurements for every single phosphorylation site. As seen in this chapter, this is not possible using information dependent acquisition. Chapter V in this thesis describes our work in developing a quantitative and **reproducible** method in order to generate robust mass spectrometry data for mathematical modeling and computational analysis.

IV.5 REFERENCES

Avraham HK, Lee TH, Koh Y, Kim TA, Jiang S, Sussman M, Samarel AM, Avraham S (2003) Vascular endothelial growth factor regulates focal adhesion assembly in human brain microvascular endothelial cells through activation of the focal adhesion kinase and related adhesion focal tyrosine kinase. *J Biol Chem* 278: 36661-36668.

Band V, Sager R (1989) Distinctive traits of normal and tumor-derived human mammary epithelial cells expressed in a medium that supports long-term growth of both cell types. *Proc Natl Acad Sci U S A* 86: 1249-1253.

Batzer AG, Rotin D, Urena JM, Skolnik EY, Schlessinger J (1994) Hierarchy of binding sites for Grb2 and Shc on the epidermal growth factor receptor. *Mol Cell Biol* 14: 5192-5201.

Cary LA, Guan JL (1999) Focal adhesion kinase in integrin-mediated signaling. *Front Biosci* 4: D102-113.

Cheng N, Brantley DM, Liu H, Lin Q, Enriquez M, Gale N, Yancopoulos G, Cerretti DP, Daniel TO, Chen J (2002) Blockade of EphA receptor tyrosine kinase activation inhibits vascular endothelial cell growth factor-induced angiogenesis. *Mol Cancer Res* 1: 2-11.

Chung CY, Murphy-Ullrich JE, Erickson HP (1996) Mitogenesis, cell migration, and loss of focal adhesions induced by tenascin-C interacting with its cell surface receptor, annexin II. *Mol Biol Cell* 7: 883-892.

Cobb MH, Hepler JE, Cheng M, Robbins D (1994) The mitogen-activated protein kinases, ERK1 and ERK2. *Semin Cancer Biol* 5: 261-268.

Coggeshall KM (1998) Inhibitory signaling by B cell Fc gamma RIIB. *Curr Opin Immunol* 10: 306-312.

D'Souza B, Taylor-Papadimitriou J (1994) Overexpression of ERBB2 in human mammary epithelial cells signals inhibition of transcription of the E-cadherin gene. *Proc Natl Acad Sci U S A* 91: 7202-7206.

Davis MA, Ireton RC, Reynolds AB (2003) A core function for p120-catenin in cadherin turnover. *J Cell Biol* 163: 525-534.

Dittmar T, Husemann A, Schewe Y, Nofer JR, Niggemann B, Zanker KS, Brandt BH (2002) Induction of cancer cell migration by epidermal growth factor is initiated by specific phosphorylation of tyrosine 1248 of c-erbB-2 receptor via EGFR. *Faseb J* 16: 1823-1825.

Eriksson L, Johansson E, Kettaneh-Wold N, Wold S (2001) *Multi- and Megavariate Data Analysis Principles and Applications*. Umetrics, Umea, Sweden.

Falsey RR, Marron MT, Gunaherath GM, Shirahatti N, Mahadevan D, Gunatilaka AA, Whitesell L (2006) Actin microfilament aggregation induced by withaferin A is mediated by annexin II. *Nat Chem Biol*: Epub 2005 Dec 2011.

Ganoth D, Bornstein G, Ko TK, Larsen B, Tyers M, Pagano M, Hershko A (2001) The cell-cycle regulatory protein Cks1 is required for SCF(Skp2)-mediated ubiquitinylation of p27. *Nat Cell Biol* 3: 321-324.

Gaudet S, Janes KA, Albeck JG, Pace EA, Lauffenburger DA, Sorger PK (2005) A Compendium of Signals and Responses Triggered by Prodeath and Prosurvival Cytokines. *Mol Cell Proteomics* 4: 1569-1590.

Gensler M, Buschbeck M, Ullrich A (2004) Negative regulation of HER2 signaling by the PEST-type protein-tyrosine phosphatase BDP1. *J Biol Chem* 279: 12110-12116.

Grovdal LM, Stang E, Sorkin A, Madshus IH (2004) Direct interaction of Cbl with pTyr 1045 of the EGF receptor (EGFR) is required to sort the EGFR to lysosomes for degradation. *Exp Cell Res* 300: 388-395.

Hautaniemi S, Yli-Harja O, Astola J, Kauraniemi P, Kallioniemi A, Wolf M, Ruiz J, Mousses S, Kallioniemi O-P (2003) Analysis and Visualization of Gene Expression Microarray Data in Human Cancer Using Self-Organizing Maps. *Machine Learning* 52: 45-66.

Hendriks BS, Opresko LK, Wiley HS, Lauffenburger D (2003a) Coregulation of epidermal growth factor receptor/human epidermal growth factor receptor 2

(HER2) levels and locations: quantitative analysis of HER2 overexpression effects. *Cancer Res* 63: 1130-1137.

Hendriks BS, Opresko LK, Wiley HS, Lauffenburger D (2003b) Quantitative analysis of HER2-mediated effects on HER2 and epidermal growth factor receptor endocytosis: distribution of homo- and heterodimers depends on relative HER2 levels. *J Biol Chem* 278: 23343-23351.

Hendriks BS, Orr G, Wells A, Wiley HS, Lauffenburger DA (2005) Parsing ERK activation reveals quantitatively equivalent contributions from epidermal growth factor receptor and HER2 in human mammary epithelial cells. *J Biol Chem* 280: 6157-6169.

Holmqvist K, Cross M, Riley D, Welsh M (2003) The Shb adaptor protein causes Src-dependent cell spreading and activation of focal adhesion kinase in murine brain endothelial cells. *Cell Signal* 15: 171-179.

Janes KA, Albeck JG, Gaudet S, Sorger PK, Lauffenburger DA, Yaffe MB (2005) A systems model of signaling identifies a molecular basis set for cytokine-induced apoptosis. *Science* 310: 1646-1653.

Janes KA, Kelly JR, Gaudet S, Albeck JG, Sorger PK, Lauffenburger DA (2004) Cue-signal-response analysis of TNF-induced apoptosis by partial least squares regression of dynamic multivariate data. *J Comput Biol* 11: 544-561.

Jawhari AU, Farthing MJ, Pignatelli M (1999) The E-cadherin/epidermal growth factor receptor interaction: a hypothesis of reciprocal and reversible control of intercellular adhesion and cell proliferation. *J Pathol* 187: 155-157.

Kinch MS, Carles-Kinch K (2003) Overexpression and functional alterations of the EphA2 tyrosine kinase in cancer. *Clin Exp Metastasis* 20: 59-68.

Knowlden JM, Hutcheson IR, Jones HE, Madden T, Gee JM, Harper ME, Barrow D, Wakeling AE, Nicholson RI (2003) Elevated levels of epidermal growth factor receptor/c-erbB2 heterodimers mediate an autocrine growth regulatory pathway in tamoxifen-resistant MCF-7 cells. *Endocrinology* 144: 1032-1044.

Kohonen T (2001) *Self-Organizing Maps*. Springer, Berlin, Germany.

Lin Q, Lo CG, Cerione RA, Yang W (2002) The Cdc42 target ACK2 interacts with sorting nexin 9 (SH3PX1) to regulate epidermal growth factor receptor degradation. *J Biol Chem* 277: 10134-10138.

Liu JW, Shen JJ, Tanzillo-Swarts A, Bhatia B, Maldonado CM, Person MD, Lau SS, Tang DG (2003) Annexin II expression is reduced or lost in prostate cancer cells and its re-expression inhibits prostate cancer cell migration. *Oncogene* 22: 1475-1485.

Lorch JH, Klessner J, Park JK, Getsios S, Wu YL, Stack MS, Green KJ (2004) Epidermal growth factor receptor inhibition promotes desmosome assembly and strengthens intercellular adhesion in squamous cell carcinoma cells. *J Biol Chem* 279: 37191-37200.

Manabe R, Kovalenko M, Webb DJ, Horwitz AR (2002) GIT1 functions in a motile, multi-molecular signaling complex that regulates protrusive activity and cell migration. *J Cell Sci* 115: 1497-1510.

Mariner DJ, Davis MA, Reynolds AB (2004) EGFR signaling to p120-catenin through phosphorylation at Y228. *J Cell Sci* 117: 1339-1350.

Moon HS, Choi EA, Park HY, Choi JY, Chung HW, Kim JI, Park WI (2001) Expression and tyrosine phosphorylation of E-cadherin, beta- and gamma-catenin, and epidermal growth factor receptor in cervical cancer cells. *Gynecol Oncol* 81: 355-359.

Muraille E, Pesesse X, Kuntz C, Erneux C (1999) Distribution of the src-homology-2-domain-containing inositol 5-phosphatase SHIP-2 in both non-haemopoietic and haemopoietic cells and possible involvement of SHIP-2 in negative signalling of B-cells. *Biochem J* 342 Pt 3: 697-705.

Okabayashi Y, Kido Y, Okutani T, Sugimoto Y, Sakaguchi K, Kasuga M (1994) Tyrosines 1148 and 1173 of activated human epidermal growth factor receptors are binding sites of Shc in intact cells. *J Biol Chem* 269: 18674-18678.

Playford MP, Schaller MD (2004) The interplay between Src and integrins in normal and tumor biology. *Oncogene* 23: 7928-7946.

Pradhan M, Coggeshall KM (1997) Activation-induced bi-dentate interaction of SHIP and Shc in B lymphocytes. *J Cell Biochem* 67: 32-42.

Prager GW, Breuss JM, Steurer S, Olcaydu D, Mihaly J, Brunner PM, Stockinger H, Binder BR (2004) Vascular endothelial growth factor receptor-2-induced initial endothelial cell migration depends on the presence of the urokinase receptor. *Circ Res*: Epub 2004 May 2006.

Qu CK, Yu WM, Azzarelli B, Feng GS (1999) Genetic evidence that Shp-2 tyrosine phosphatase is a signal enhancer of the epidermal growth factor receptor in mammals. *Proc Natl Acad Sci U S A* 96: 8528-8533.

Riggins RB, Quilliam LA, Bouton AH (2003) Synergistic promotion of c-Src activation and cell migration by Cas and AND-34/BCAR3. *J Biol Chem* 278: 28264-28273.

Roskoski R, Jr. (2004) Src protein-tyrosine kinase structure and regulation. *Biochem Biophys Res Commun* 324: 1155-1164.

Rothhut B (1997) Participation of annexins in protein phosphorylation. *Cell Mol Life Sci* 53: 522-526.

Salcini AE, Chen H, Iannolo G, De Camilli P, Di Fiore PP (1999) Epidermal growth factor pathway substrate 15, Eps15. *Int J Biochem Cell Biol* 31: 805-809.

Sanchez-Pulido L, Martin-Belmonte F, Valencia A, Alonso MA (2002) MARVEL: a conserved domain involved in membrane apposition events. *Trends Biochem Sci* 27: 599-601.

Satoh T, Kato J, Nishida K, Kaziro Y (1996) Tyrosine phosphorylation of ACK in response to temperature shift-down, hyperosmotic shock, and epidermal growth factor stimulation. *FEBS Lett* 386: 230-234.

Schulze WX, Deng L, Mann M (2005) Phosphotyrosine interactome of the ErbB-receptor kinase family. *Mol Syst Biol* 1: doi:10.1038/msb4100012.

Semino CE, Kamm RD, Lauffenburger DA (2006) Autocrine EGF receptor activation mediates endothelial cell migration and vascular morphogenesis induced by VEGF under interstitial flow. *Exp Cell Res*: Epub 2005 Dec 2007.

Slotky M, Shapira M, Ben-Izhak O, Linn S, Futerman B, Tsalic M, Herskho DD (2005) The expression of the ubiquitin ligase subunit Cks1 in human breast cancer. *Breast Cancer Res* 7: R737-744.

Sorkin A, Mazzotti M, Sorkina T, Scotto L, Beguinot L (1996) Epidermal growth factor receptor interaction with clathrin adaptors is mediated by the Tyr974-containing internalization motif. *J Biol Chem* 271: 13377-13384.

Soulet F, Yarar D, Leonard M, Schmid SL (2005) SNX9 regulates dynamin assembly and is required for efficient clathrin-mediated endocytosis. *Mol Biol Cell* 16: 2058-2067.

Spencer KS, Graus-Porta D, Leng J, Hynes NE, Klemke RL (2000) ErbB2 is necessary for induction of carcinoma cell invasion by ErbB family receptor tyrosine kinases. *J Cell Biol* 148: 385-397.

Stampfer MR, Bartley JC (1985) Induction of transformation and continuous cell lines from normal human mammary epithelial cells after exposure to benzo[a]pyrene. *Proc Natl Acad Sci U S A* 82: 2394-2398.

Taya S, Inagaki N, Sengiku H, Makino H, Iwamatsu A, Urakawa I, Nagao K, Kataoka S, Kaibuchi K (2001) Direct interaction of insulin-like growth factor-1 receptor with leukemia-associated RhoGEF. *J Cell Biol* 155: 809-820.

Taylor V, Wong M, Brandts C, Reilly L, Dean NM, Cowser LM, Moodie S, Stokoe D (2000) 5' phospholipid phosphatase SHIP-2 causes protein kinase B inactivation and cell cycle arrest in glioblastoma cells. *Mol Cell Biol* 20: 6860-6871.

Tsubouchi A, Sakakura J, Yagi R, Mazaki Y, Schaefer E, Yano H, Sabe H (2002) Localized suppression of RhoA activity by Tyr31/118-phosphorylated paxillin in cell adhesion and migration. *J Cell Biol* 159: 673-683.

Ultsch A, Siemon H (1989) Exploratory data analysis: Using Kohonen networks on transputers.

Vesanto J, Himberg J, Alhoniemi E, Parkankangas J (2000) SOM toolbox for Matlab 5. In *Technical Report A57*. Helsinki University of Technology, Helsinki, Finland.

Vulin AI, Jacob KK, Stanley FM (2005) Integrin activates receptor-like protein tyrosine phosphatase alpha, Src, and Rho to increase prolactin gene expression through a final phosphatidylinositol 3-kinase/cytoskeletal pathway that is additive with insulin. *Endocrinology* 146: 3535-3546.

Wang FM, Liu HQ, Liu SR, Tang SP, Yang L, Feng GS (2005a) SHP-2 promoting migration and metastasis of MCF-7 with loss of E-cadherin, dephosphorylation of FAK and secretion of MMP-9 induced by IL-1beta in vivo and in vitro. *Breast Cancer Res Treat* 89: 5-14.

Wang Y, Hailey J, Williams D, Wang Y, Lipari P, Malkowski M, Wang X, Xie L, Li G, Saha D, Ling WL, Cannon-Carlson S, Greenberg R, Ramos RA, Shields R, Presta L, Brams P, Bishop WR, Pachter JA (2005b) Inhibition of insulin-like growth factor-I receptor (IGF-IR) signaling and tumor cell growth by a fully human neutralizing anti-IGF-IR antibody. *Mol Cancer Ther* 4: 1214-1221.

Webb DJ, Donais K, Whitmore LA, Thomas SM, Turner CE, Parsons JT, Horwitz AF (2004) FAK-Src signalling through paxillin, ERK and MLCK regulates adhesion disassembly. *Nat Cell Biol* 6: 154-161.

Xia L, Wang L, Chung AS, Ivanov SS, Ling MY, Dragoi AM, Platt A, Gilmer TM, Fu XY, Chin YE (2002) Identification of both positive and negative domains within the epidermal growth factor receptor COOH-terminal region for signal transducer and activator of transcription (STAT) activation. *J Biol Chem* 277: 30716-30723.

Yamboliev IA, Chen J, Gerthoffer WT (2001) PI 3-kinases and Src kinases regulate spreading and migration of cultured VSMCs. *Am J Physiol Cell Physiol* 281: C709-718.

Yang CM, Lin MI, Hsieh HL, Sun CC, Ma YH, Hsiao LD (2005) Bradykinin-induced p42/p44 MAPK phosphorylation and cell proliferation via Src, EGF receptors, and PI3-K/Akt in vascular smooth muscle cells. *J Cell Physiol* 203: 538-546.

Yarden Y, Sliwkowski MX (2001) Untangling the ErbB signalling network. *Nat Rev Mol Cell Biol* 2: 127-137.

Zhang X, Lin M, van Golen KL, Yoshioka K, Itoh K, Yee D (2005a) Multiple Signaling Pathways are Activated During Insulin-like Growth Factor-I (IGF-I) Stimulated Breast Cancer Cell Migration. *Breast Cancer Res Treat* 93: 159-168.

Zhang Y, Wolf-Yadlin A, Ross PL, Pappin DJ, Rush J, Lauffenburger DA, White FM (2005b) Time-resolved Mass Spectrometry of Tyrosine Phosphorylation Sites in the Epidermal Growth Factor Receptor Signaling Network Reveals Dynamic Modules. *Mol Cell Proteomics* 4: 1240-1250.

**V MULTIPLE REACTION MONITORING AND
INFORMATION DEPENDENT ACQUISITION:
TOWARDS A REPRODUCIBLE, QUANTITATIVE AND
ROBUST MASS SPECTROMETRY BASED
TECHNOLOGY FOR PROTEOMICS**

V.1 SUMMARY

One basic premise to obtain data amenable to computational analysis and mathematical modeling is that the data needs to be quantitative, reproducible and significant from a statistic point of view. The results from the previous chapter showed that while our mass spectrometry based technology for phosphoproteomics is highly quantitative, phosphorylation site specific and the data obtained by it can be used for mathematical modeling and computational analysis its reproducibility is still very low. This results in the heterogeneous detection of many phosphorylation sites across many mass spectrometry runs with small overlap among them, in fact in the previous chapter out of a total of 332 phosphorylated peptides found only 68 (20.4%) were found across all conditions, notwithstanding the copious amount of new insights we gain from the analysis of just this subset peptides we can not but wonder how much information and how much deeper insight we could have gained if we had had a more robust mass spectrometry technology allowing for a significantly higher peptide detection overlap across all conditions .

In this chapter I will talk about an expansion to our technology that we have developed to specifically tackle the issue of mass spectrometry run-to-run reproducibility. To solve this issue we complemented our information dependent acquisition (IDA) methodology with multiple reaction monitoring (MRM) based mass spectrometry. MRM has been largely use as a mass spectrometry technique when it is desired to find a particular chemical in a sample, for instance

MRM is a preferred tool for doping control in sports competitions and forensic analysis. However applications of MRM to the field of phosphoproteomics are few.

To complement our current IDA based technology we have designed, tested, validated and implemented the MRM methodology for our phosphoproteomics studies. To construct our MRM methodology we took as input phosphorylated peptides detected in IDA mode – characterized by their LC elution profile, parent ion m/z , one b and one y ion fragment and iTRAQ markers. In MRM mode the mass spectrometer is programmed to search for these peptides (according to their LC elution order) and a peptide is considered to be found when both b and y ions originating from a given parent ion are co-detected.

To test the application of MRM to phosphoproteomics and to gain further dynamic resolution on the activation of the EGFR network we have designed what is to our knowledge the largest MRM experiment ever done. Out of a search space of 226 peptides we were able to find 223 by MRM of which 199 (89.2%) were found in every single mass spectrometry run. In contrast using IDA only 186 peptides were detected of which just 63 (33.8%) were found in every single mass spectrometry run. This highlights the enormous contribution of MRM to build a robust and reproducible mass spectrometry based phosphoproteomics technology.

Finally, the successful completion of our experiment allowed us to characterize with high temporal resolution (seven time points, spread across 32 minutes) the phosphorylation of 222 phosphopeptides – covering 200

phosphotyrosine sites (34 novel), 6 phosphoserine sites (1 novel) and 2 phosphothreonine sites in 143 proteins – after EGF stimulation of human mammary epithelial cells (HMECs) (Stampfer & Bartley, 1985).

The data presented in this chapter details the process of building, validating and characterizing MRM for phosphoproteomics. We are currently working on a manuscript to publish this data in a peer reviewed journal.

V.2 MATERIALS AND METHODS

V.2.1 Cell Culture and Stimulation

184A1 HMECs (human mammary epithelial cells (Stampfer & Bartley, 1985)) were a kind gift from Martha Stampfer (Lawrence Berkeley Laboratory, Berkeley CA) via Steve Wiley (Pacific Northwest National Laboratory, Richland WA) and were maintained in DFCI-1 medium supplemented with 12.5 ng/ml EGF, as in (Band & Sager, 1989). Cells were washed with PBS and incubated for 12 hours in serum free media (DFCI-1 without EGF, bovine pituitary extract, or fetal bovine serum) after 80% confluence was reached in 15 cm plates ($\sim 10^7$ cells). Synchronized cells were then stimulated with 100 ng/ml EGF in serum free media for 1, 2, 4, 8, 16 or 32 minutes, or left untreated with serum free media for 4 min as control before lysing. Because the 4 minutes time point was used as time course normalization control 2 plates in stead of 1 were stimulated for 4 minutes in every single time course.

V.2.2 Protein Quantification

Protein concentration of cell lysates was determined using the micro bicinchoninic acid assay (micro-BSA – Pierce) according to the manufacturer's protocol.

V.2.3 Sample Preparation

Cells were lysed with 8M urea / 1mM sodium orthovanadate. After DTT reduction and iodoacetamide alkylation proteins were digested with trypsin. Tryptic peptides were desalted and fractionated on a C18 Sep-Pak Plus cartridge (Waters), and the 25% MeCN fraction was divided into 10 equivalent aliquots which were lyophilized to dryness. One aliquot of sample from each condition was labeled with one tube of iTRAQ reagent (Applied Biosystems, CA) (following manufacturer's directions). Samples labeled with four different isoforms of the iTRAQ reagent were combined, dried completely, and saved at -80°C. This process was repeated to generate two biological and two analytical duplicate sets of time courses each divided into two samples (sub time courses, 0, 1, 2 and 4 minutes and 4, 8, 16, and 32 minutes for labeling with iTRAQ 114, 115, 116 and 117 tags respectively).

V.2.4 Peptide IP

10 µg of protein G Plus-agarose beads (Calbiochem) were incubated with 12 µg of each anti-phosphotyrosine antibody (pTyr100 (Cell Signaling Technology), PT-66(Sigma)) in 200 µl of IP buffer for 8 hr at 4°C. The beads were rinsed once with 400 µl of IP buffer at 4°C. iTRAQ labeled sample were dissolved in 150 µl of IP buffer (100 mM Tris, 100 mM NaCl, 1% NP-40, pH 7.4)

and 300ul of water. After pH was adjusted to 7.4 with 0.5 M Tris buffer pH 8.5, the sample was mixed with the antibody bound protein G Plus-agarose beads, and was incubated overnight at 4°C. The protein G Plus-agarose beads were spun down for 5 minutes at 6000 rpm and the supernatant was separated and saved. Antibody-bound beads were washed once with 400 µl IP buffer for 10 minutes and twice with rinse buffer (100 mM Tris, 100 mM NaCl, pH 7.4) for 5 minutes at 4°C. Phosphotyrosine containing peptides were eluted from antibody with 70 µl of 100 mM glycine pH 2.5 for 30 min at room temperature.

V.2.5 IMAC and IDA Mass Spectrometry Analysis

Phosphopeptide enrichment on IMAC was performed as described (Zhang et al., 2005c). Peptides retained on the IMAC column were eluted to a C₁₈ capillary precolumn (100 µm ID, 360 µm OD) with 50 µl of 250 mM Na₂HPO₄, pH 8.0. After 10 min rinse (0.1% HOAc), the precolumn was connected to an analytical capillary column with an integrated electrospray tip (~1 µm orifice). Peptides were eluted using a 100 min gradient with solvent A (H₂O/HOAc, 99/1 vol/vol) and B (H₂O/MeCN/HOAc, 29/70/1 vol/vol): 10 min from 0% to 15% B, 75 min from 15% to 40% B, 15 min from 40% to 70% B. Eluted peptides were directly electrosprayed into a quadrupole time-of-flight mass spectrometer (QSTAR XL Pro, Applied Biosystems). MS/MS spectra of four or five most intense peaks with charges 2 to 5 in the MS scan were automatically acquired in

information-dependent acquisition with previously selected peaks excluded for 40 seconds. Electrospray voltage was set to 1800 V and curtain gas at 20 nominal units. MS/MS/MS experiments were run in the positive mode with MS TOF masses filter set between 370 to 1500 a.m.u and MS accumulation times of 2.5 seconds. In MS/MS mode accumulation times were set at 3.5 second for the 3 most abundant ions and 4 seconds for the following 2. Q2 transmissions was set to spend 40% of MS/MS time between 80 and 150 a.m.u for clear iTRAQ quantification, 15% between 150 and 300 a.m.u for detection of phosphotyrosine immonium ion, 20% between 300 and 600 a.m.u and 25% between 600 and 1500 a.m.u. to focus on peptide identification. Collision energies (CE) were calculated as a linear function of m/z , with the slope and offset being empirical functions of the charge. For b and y fragment ions the parameter used were: slope = 0.0630, 0.0634, 0.0614 and 0.0518; offset = -5.5, -3, -1 and 1 for precursor ions of charge 2, 3, 4 and 5 respectively.

V.2.6 MRM Mass Spectrometry Analysis

Sample preparation, peptide IP, IMAC, C18 pre-columns, C18 analytical columns and liquid chromatography gradient were the same as above. Eluted peptides were directly electrosprayed into a triple quadrupole linear ion trap mass spectrometer (QTRAP 4000, Applied Biosystems). Electrospray was set at 2000 V and curtain gas at 20 nominal units. Enhanced MS (EMS) and multiple reaction

monitoring (MRM) were run in the positive mode with Q0 trapping activated. In EMS mode masses between 370 and 1500 a.m.u at scan rate of 1000 a.m.u per cycle were scanned. For MRM mode Q1 resolution was set to high and Q3 resolution was set to unit. Dwelling time for all product ions was set to 35 msec and CE were calculated as above, with the correction that to ensure fragmentation at low m/z for iTRAQ quantification CE settings for iTRAQ ions, were calculated with a 25% slope increase and a 6 units higher offset. CE for b and y ions was calculated using the same parameters as above. To ensure fragmentation at low m/z for iTRAQ quantification in the QTRAP CE setting for iTRAQ ions, were calculated with a 25% slope increase and a 5 units higher offset.

V.2.7 Phosphopeptide Sequencing, Data Clustering and Analysis

For IDA experiments, MS/MS spectra were extracted and searched against human protein database (NCBI) using ProQuant (Applied Biosystems) as described by the manufacturer. An interrogator database was generated by predigesting the human protein database with trypsin and allowing one miscleavage and up to six modifications on a single peptide (phosphotyrosine \leq 2, phosphoserine \leq 1, phosphothreonine \leq 1, iTRAQ-lysine \leq 4, and iTRAQ-tyrosine \leq 4). Mass tolerance was set to 2.2 amu for precursor ions and 0.15 amu for fragment ions. Grouping of MS/MS spectra from different cycles was set to

zero. Phosphotyrosine-containing peptides identified by ProQuant were manually validated. ProQuant quantification results were corrected by removing contaminant signals near iTRAQ tag peaks. Data was further corrected with values generated from the peak areas of non-phosphorylated peptides to account for possible variations in the starting amounts of sample for each time point. All the data from each analysis was normalized by the 4 min peak area values for integration of sub time courses.

For MRM experiments, peak detection and automatic quantification methods were built for each method using Analyst 1.4.1 (Applied Biosystems). Bunching factor and number of smooths, were set to 1. Peak detection was inspected visually, to ensure correct peak identification (co-elution) and shape (smoothness). For quantifying peptides eluting at the very beginning or end of any MRM methods, cases in which only the rising or decreasing section of a peak was detected the manual quantification option of Analyst 1.4.1 was used. As an alternative for boundary peak quantification and also used to quantify peptides with noisy iTRAQ signal, peptides were quantified directly from the MRM raw data in the original acquisition wiff file. Intensity counts of 3 – 5 iTRAQ analytical replicates were taken around the peptide b and y ions mean time of co-elution and averaged for quantification. For peptides with good elution patterns wiff file quantification closely reproduced the results obtained by automatic quantification. As in IDA data was corrected using the peak areas of non-phosphorylated peptides for possible variations in the starting amounts of sample

for each time point. For consistency normalization runs were also done on the EMS/MRM mode.

All the data from each analysis, for both IDA and MRM analysis was normalized by the 4 min peak area values for integration of the sub time courses into a full time course.

V.3 RESULTS

V.3.1 A Robust Technology for Mass Spectrometry Based Phosphoproteomics

The cell signaling discovery and monitoring of phosphorylation (CSDMP) technology is an extension of our previously developed LCMS/MS technology for characterization and quantification of tyrosine phosphorylation events in cell signaling networks (Zhang et al., 2005c) (Figure V.1). Briefly, following cell lysis after the desired stimulation conditions proteins are denatured alkylated and digested with trypsin. Phosphorylated peptides are extracted by reverse phase chromatography and aliquoted. Four aliquots from four different samples are labeled using each one of four different iTRAQ tags – if there is more than four samples to be analyzed the samples are divided in sets of four and one of the samples is included in every single set to allow for direct comparison across groups –. After labeling, the samples in one set are mixed, immunoprecipitated using anti-phosphotyrosine antibody and further purified by IMAC enrichment. The pulled down peptides are analyzed by LC-ESI-MS/MS using a quadrupole – time of flight mass spectrometer (QSTAR-XL) set in information dependent acquisition (IDA) mode. Our group has successfully applied this technology to EGFR family members and T-Cell and insulin mediated signaling (Kim & White, 2006; Schmelzle et al., 2006; Zhang et al., 2005c)MSB.

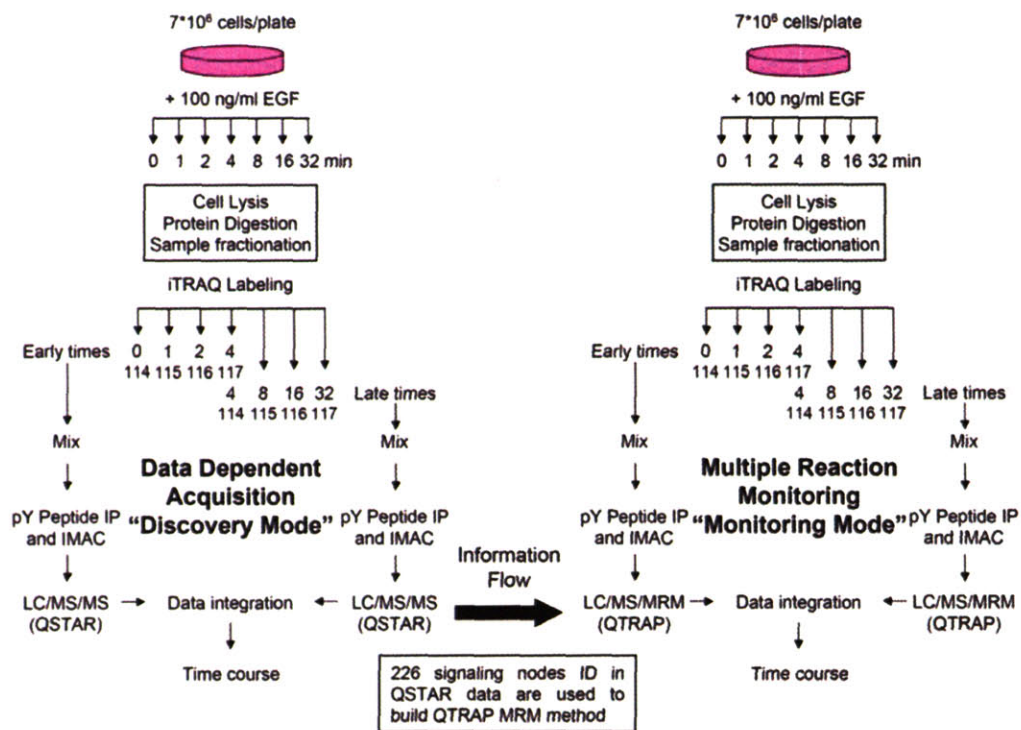


Figure V.1 – Schematics of CSDMP Method. Biological duplicates of eight 80 to 90% confluent plates of parental HMECs were treated with 100 ng/ml EGF for 0, 1, 2, 4 (2x), 8, 16 and 32 minutes respectively after 12 hour serum starvation. Cells were lysed with 8M urea / 1mM sodium orthovanadate. After DTT reduction and iodoacetamide alkylation proteins were digested with trypsin. Tryptic peptides were desalted and fractionated on a C18 Sep-Pak Plus cartridge (Waters), and the 25% MeCN fraction was divided into 10 equivalent aliquots which were lyophilized to dryness. One aliquot of sample from each condition was labeled with one tube of iTRAQ reagent (Applied Biosystems, CA) (following manufacturer's directions). Samples labeled with four different isoforms of the iTRAQ reagent were combined, immunoprecipitated and analyzed by LC-MS-MS/MS and LC-EMS-MRM respectively. This process was repeated to generate two biological and two analytical duplicate sets of time courses each divided into two samples (sub time courses, 0, 1, 2 and 4 minutes and 4, 8, 16, and 32 minutes for labeling with iTRAQ 114, 115, 116 and 117 tags respectively). The MRM method was designed based on several previous IDA analysis of EGFR mediated phosphorylation in parental HMECs and tool into account LC-elution profile of the peptides.

The main problem that we have found when applying this technology is based on the nature of IDA methods – where for each MS scan the top n most abundant peptides are chosen for MS/MS (with n between 1 and 10) – resulting in that any change in the sample nature translates into the identification and quantification of a different set of tyrosine phosphorylated peptides in each

analysis with only a subset of the peptides being identified in every single run. As corollary, the percentage of peptides that can be found in biological replicates, different sections of a time course or across different biological condition to perform quantitative analysis of statistically significant data decreases as the number of replicates, time points or biological conditions increase.

To solve the aforementioned issue we had expanded our technology to go one step further from discovery/IDA mode into monitoring/MRM mode (Figure V.1). Our CSDMP technology builds on our previous methodology by combining the results of various QSTAR IDA experiments to generate a multiple reaction monitoring (MRM) method that can be implemented using a triple quadrupole mass spectrometer (QTRAP 4000).

To build our MRM method we use four parameters that are readily obtained from one or multiple IDA experiments: 1) Relative elution time of a peptide with respect to others, 2) intact peptides (parent ions) charge state (z) and m/z , 3) characteristic b ion and y ion m/z for a given parent ion and 4) collision energy required for sequence (b and y ions) and quantification (iTRAQ markers) related fragmentation. Figure V.2 depicts the process of building and MRM method for the tryptic peptide containing EGFR pY1173. The peptide elution profile in all IDA runs that are going to be used to build the method is characterized and a list of all peptides eluting before and after it is constructed. This list allows us to determine when in the MRM experiment should we look for the peptide (the list is not necessary if less than 300 fragments are being detected). At the same time, parent ion and characteristic b and y ions m/z are

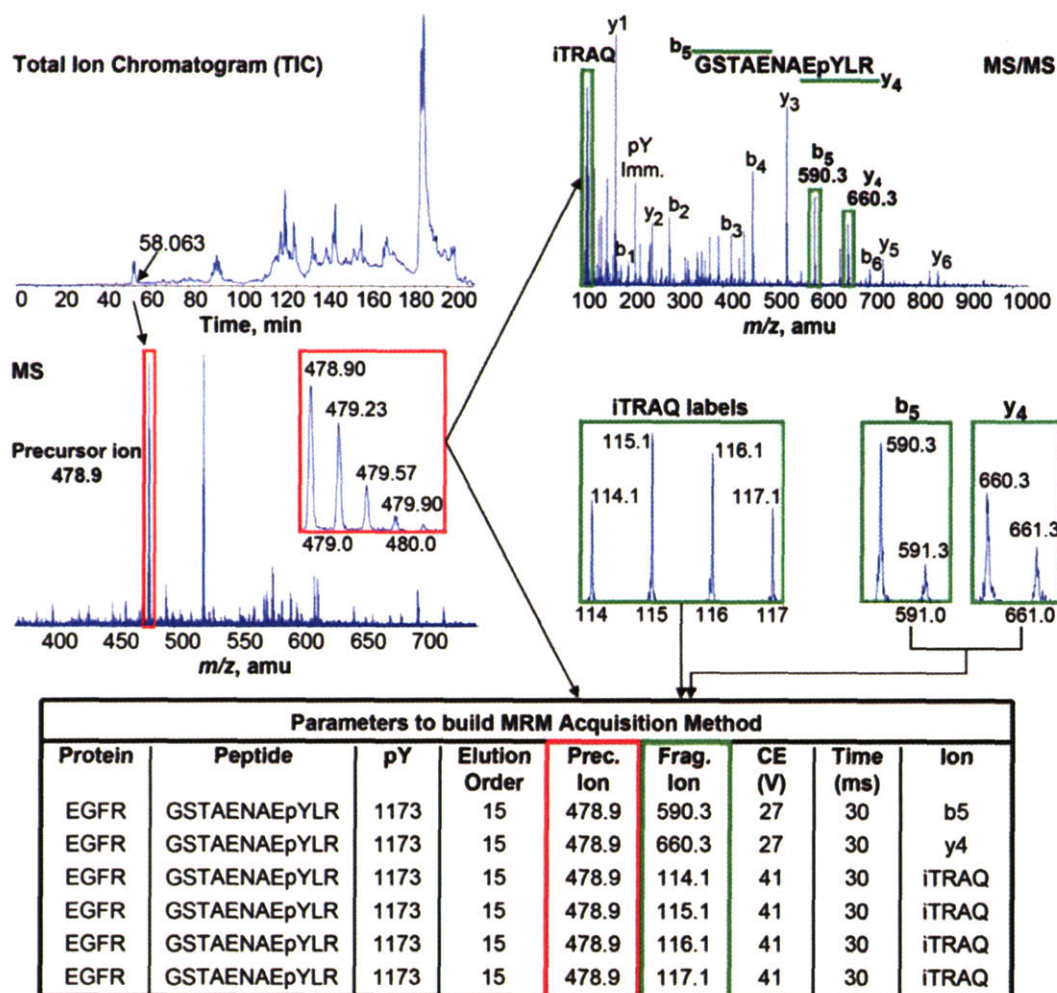


Figure V.2 – How to construct MRM method from IDA data. Elution time of the peptides is obtained from total ion chromatogram, parent ion m/z from the corresponding MS scan and b and y m/z ions from the appropriate MS/MS scan. Peptides are ordered according to their LC elution profiles and CEs are calculated as a linear function of parent ion charge. To favor marker ion detection iTRAQ CE is calculated with a 25% increase in slope and an increase of 6 eV in the baseline with respect to b and y ions. Accumulation times can also be adjusted, but in this work we use 35 ms for all selected fragments. The table shows the necessary information to detect EGFR pY1173 in MRM mode, including its expected relative elution order with respect to other sought peptides. Elution order is important only when monitoring more than 300 fragment ions (i.e. 150 peptides without quantification or 50 peptides including quantification).

identified from an IDA obtained MS/MS spectra. b and y ions are selected according to their signal intensity in MS/MS with the constraints that each of them should be at least 4 residues long and their added length should be at least 9

amino acids, unless the peptide is shorter than this length. Because ions with charge 2 and 3 were found to fly better in the QTRAP than ions with charge 4 or 5 in many cases it was necessary to convert the parent ion m/z values found in IDA to their equivalent values at charge 2 or 3, with the restriction that all parent ions m/z should be lower than 1200 a.m.u.

The collision energy both in IDA and MRM methods is a linear function of the parent ion m/z , with the slope and offset of the equation being an empirical function of the parent ion charge state. The collision energy for detection is calculated keeping the parameters that were used in the IDA experiments whereas to ensure iTRAQ fragmentation the collision energy for quantification is calculated by increasing the slope of the equation 25% and the offset 6 units. Finally a table like the one shown at the bottom of Figure V.2 for EGFR pY1173 can be obtained for every peptide of interest and an MRM method is constructed by merging these tables together in order from first to last eluting peptide.

It is important to note that the QTRAP can monitor a maximum of 300 reaction fragments per MRM method, taking into consideration that for each peptide we wanted to detect we needed to track six fragment ions (b and y ions plus four iTRAQ markers). Given that we were looking for 226 peptides, including negative controls, which gave us a total 1356 fragments to look, we clearly exceeded the capabilities of using a single MRM method. To monitor all these ions, for a single LC run we needed to run five sequential MRM methods keeping into consideration the peptide elution order. This is the reason why it is of the utmost importance to generate a correct list for peptides relative time elution as

any mistake on the order of this list might redound in looking for a peptide at the wrong time and hence not finding it.

In summary our CSDMP technology extend on our previous phosphoproteomic technique by using the IDA generated data to construct an MRM technique to reproducibly monitor for the presence of a list of peptides of interest across various samples, time points or biological conditions.

V.3.2 QTRAP/CSDMP Data Output

The same software that is used for data acquisition (Analyst™ 1.4.1) was used to process and quantify the results. We have applied our new CSDMP technology to identify and quantify 223 peptides, covering 208 phosphorylation sites (35 novel) in 143 proteins during a 7 time points time course (0, 1, 2, 4, 8, 16, 32 minutes) of EGF (100 ng/ml) stimulated human mammary epithelial cells (HMECs) (Stampfer & Bartley, 1985). Due to the limitation of iTRAQ labels (it is only possible to label 4 independent samples at time) we performed four QTRAP runs in order to obtain biological replicates for our time courses.

The QTRAP data output can be processed to extract the elution profile of all the peptides that were included in the MRM methods (Figure V.3). Figure V.3A shows the elution profile of EGFR pY1173 tryptic peptide. The left column shows detection of b₅ and y₄ ion and quantification of phosphorylation for the time course early points, whereas the right column shows detection of b₅ and y₄

ions and quantification of phosphorylation for the time course late points in a second CSDMP experiment. Note that in both columns, b_5 , y_4 and iTRAQ ions co-elute which is a necessary condition to trust both identification and quantification. The data in Figure V.3A can be used to construct the time course shown in Figure V.3C. Co-elution of b and y ions is used to confirm peptide identification, while the area under the curve of the corresponding iTRAQ tags is used to quantify phosphorylation at the different time points. To link the 2 MRM runs covering the early and late section of the time course all time points are normalized to the overlapping time point, in this case 4 minutes. Error bars are from biological replicates, requiring 4 CSDMP/MRM runs to obtain 2 time courses for the same peptide.

Figures V.3B and V.3D are analogous to V.3A and V.3C in that they show the identification, quantification and time course construction (also from biological replicates) for N-WASP tryptic peptide containing pY256. In this case b_4 and y_4 ions were used for peptide identification. Figure V.3B highlights one of the key advantages of our method: the ability to distinguish between real identification events and false positives. By requiring co-elution of b ions which are representative of a peptide N-terminus and y ions which are representative of a peptide C-terminus we can ensure that the right peptide is identified and quantified, even in the case that two peptides with similar parent ions share homology at one of their ends. This is exactly what we see in Figure V.3B where there are 2 peaks co-eluting in all profiles presented, except for the b_4 ion which only shows one peak co-eluting with the earlier peak of the other fragment ions.

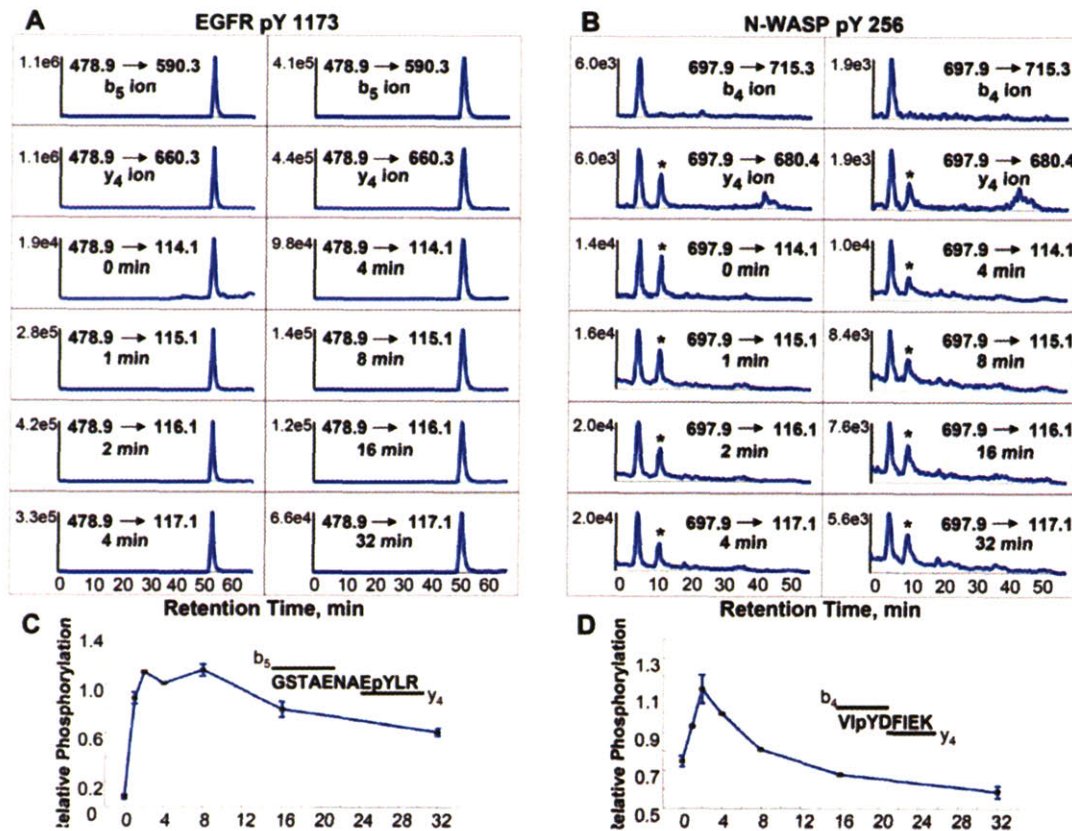


Figure V.3 – Detecting and quantifying protein phosphorylation by MRM. MRM elution profiles of EGFR pY1173 (A) and N-WASP pY256 (B) and respective phosphorylation time courses (C) and (D). Figure 3a left column shows co-elution of the b, y and iTRAQ ions identifying EGFR pY1173 and quantifying its phosphorylation after 0, 1, 2 and 4 minutes of EGF stimulation. The right column shows co-elution of the b, y and iTRAQ ions identifying EGFR pY1173 and quantifying its phosphorylation after 4, 8, 16 and 32 minutes of EGF stimulation. Figure 3c shows EGFR pY1173 relative phosphorylation time course. The data was normalized to 4 minutes to couple together both half of the time course. Error bars are from biological replicates. Figures 3b and 3d are analogous to 3a and 3c, but identify and quantify the time course for N-WASP pY256. Note that identification of b and y ions is required to ensure correct peptide ID as a false positive (marked by *) might have been validated if only the y ion would have been used for detection N-WASP pY256. Co-elution of b and y ion peptides ensures that the peptide detected is correct as it impose constrains in both the N and C terminus sequence of the peptides.

This allows us to distinguish between the real peptide (first set of co-eluting peaks) against the false positive (second set of co-eluting peaks – marked by *) to construct the N-WASP pY256 time course in Figure V.3D.

The main issue we confronted when trying to process QTRAP data to obtain nice elution profiles came from the fact that all contaminant peptides in the sample were also iTRAQ labeled. Whether this wouldn't have been a problem in a simple sample, in a complex samples such as our originating in whole cell lysates it gave rise to somewhat high levels of iTRAQ noise over time for many parent ions. In cases where we were not able to deconvolute the time profile of iTRAQ peaks co-eluting with a given peptide's characteristic b and y ions we quantified the peptides directly using the raw data files obtained from the QTRAP. To do this we found an averaged 3 to 5 unitary quantification events for each iTRAQ marker centered on the mean time of b and y ions co-elution. We further confirmed the reliability of this quantification method by comparing its results to the standard processed data in clean peptides (data not shown). The same quantification procedure was used for peptides at the very beginning or end of MRM methods where the software was unable to perform automatic peak detection.

V.3.3 Quantification Quality

One of the main concerns of using iTRAQ labels for quantification using a triple quadrupole as opposed to the quadrupole time of flight mass spectrometer which we had previously used is their difference in sensitivity and resolution. The in comparison to 3-4 ms in QSTAR) advantages of the first one is longer scan

times (35 ms per product ion) which translate into higher sensitivity (5 attomole on column), whereas the second one has lower sensitivity (100 attomole on column), but its resolution is noticeable higher ($\sim 0.06\text{-}0.07$ a.m.u at 50% of peak height) while the QTRAP resolution is right below 1 a.m.u at 50% of peak height – the upper limit to distinguish between iTRAQ peaks – ($\sim 0.3\text{-}0.5$ a.m.u at 50% of peak height in high resolution mode and $\sim 0.5\text{-}0.7$ a.m.u at 50% of peak height at unit resolution).

To confirm the quality of quantification we compared the results from QSTAR quantification selected from 1204 iTRAQ peaks with intensities between 50 and 1500 counts (to avoid including noisy measurements or measurements where the detector might have been saturated) to the same iTRAQ peaks as detected by the QTRAP. As all the data is normalized by the 4 minute time point a total of 903 normalized data points were compared. Figure V.4A shows a very good correlation between QTRAP and QSTAR quantification with an offset of 0 and a slope of nearly 1, confirming that the quality of QTRAP quantification is comparable to the best that can be obtained using the QSTAR instrument. In fact statistically speaking the QSTAR and QTRAP data follow the same distribution (Kolmogorov-Smirnov test can not reject the hypothesis that the distributions are different with p-value of 0.5745). Figure V.4B shows the distribution of the ratio of QTRAP/QSTAR quantification for the same 903 data points. In this case the data follows a very sharp Laplace distribution with location and scaling parameters of 1.01, 0.09 with very tight 95% confidence intervals ($[1.0061, 1.0130]$ and $[0.0840, 0.0957]$) respectively.

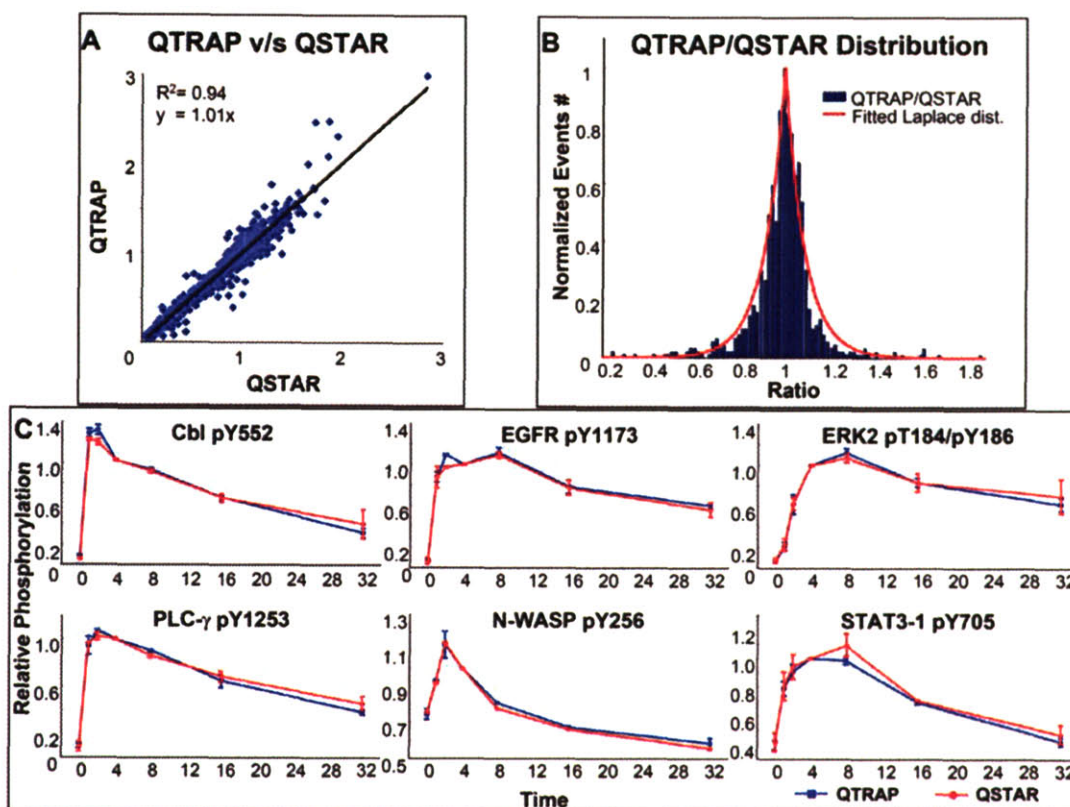


Figure V.4 – Validation of MRM quantification. (A) shows a linear regression comparing quantification results obtained by MRM/QTRAP and IDA/QSTAR methods. In order to avoid low signal to noise or detector saturation 1204 iTRAQ peaks with IDA intensities between 50 and 1500 were selected and compared to the same iTRAQ peaks detected using MRM. As all the data is normalized by the 4 minute time point a total of 903 normalized data points were used for the regression. Regression parameters show a strong correlation $R^2 = 0.94$. The slope is 1 and offset is 0 indicating that using MRM does not introduce any bias into the data and that MRM obtained quantification is as good as the best data from IDA. (B) shows the distribution of the ratio of QTRAP/QSTAR quantification for the same 903 data points. The data follows a very sharp Laplace distribution location parameter of 1.01 and scaling parameter of 0.09. Confidence intervals based on this distribution locate 85% of the QTRAP/QSTAR (MRM/IDA) ratios in the [0.84, 1.18] interval. This is consistent with the known 15-20% inherent error in measurements when using iTRAQ. (C) shows overlapping time courses for 6 phosphorylation sites (Cbl pY552, EGFR pY1173, ERK2 pT184/pY186, PLC-g pY1253, N-WASP pY256 and STAT-3 pY705) measured using QTRAP/MRM (blue) and QSTAR/IDA (red) respectively. Error bars are from biological replicates.

It is worth to note that of all the data 43 % has a QTRAP/QSTAR ratio of [0.95 1.05] and 84% has a QTRAP /QSTAR ratio of [0.85 1.15]. Of the remaining 16% where the quantification is off by more than 15% (most likely due to

sensitivity and resolution differences between the two instruments) only 14% i.e. only 2% of the total measurements are outliers where the quantification is more than 45% off. The presence of outliers and higher error peptides is readily explained by the two kind of mass spectrometer characteristics, while peptides that show low iTRAQ signal in the QSTAR (<150 counts), specially at the 0 and 32 minutes time points might be more accurately quantified using the more sensitive, higher signal to noise QTRAP, peptides that show iTRAQ shoulders, specially at 0.04 or 0.05 a.m.u lower than the 116.1 a.m.u iTRAQ label benefit from the QSTAR higher resolution which allows to distinguish between the shoulder and the actual iTRAQ peak. To address this second source of error we corrected all QTRAP quantifications by subtracting the average shoulder contribution to each iTRAQ label in every single peptide. The shoulder contribution for a given peptide was calculated by measuring both the shoulder and iTRAQ labels peak area for that peptide using the IDA data and dividing the shoulder area by the sum of shoulder and iTRAQ label. The average of all IDA hits of a given peptide was used as the QTRAP shoulder correction factor.

Figure V.4C shows the QTRAP (blue) and QSTAR (red) overlapping time courses for six different tryptic phosphopeptides from six different proteins. Note that in agreement to the previous paragraph the only measurements where the error bars do not overlap happen at 2 minutes time in Cbl pY552 and EGFR pY1173. This time point was tagged with iTRAQ 116 which is the label most prone to shoulder contamination. Other than these two measurements were the error of the mean is 5% for the former and 10% for the later the remaining 40

time points show a remarkable agreement between QTRAP and QSTAR quantification.

Overall if we consider that the internal iTRAQ quantification error is ~15% the information in this section strongly highlights the reproducibility between iTRAQ quantification using a QTRAP or QSTAR instrument.

V.3.4 Reproducibility

The main reason to develop the CSDMP technology and expand on our previous protocols was the somewhat low run to run reproducibility when using IDA methods. Adding the QTRAP/MRM methodology to our technology and using the QTRAP generated data as our final output exceeded our expectations in this area. Out of a list of 226 peptides previously found in various QSTAR runs and built into 5 sequential MRM methods 223 (98.6%) were detected using QTRAP/MRM – all of them were found in at least 2 out of 4 QTRAP runs, 217 (97.3%) were found in at least 3 runs and 199 (89.2%) were found in all the runs – (Figure V.5A). In contrast only 186 (82.3%) peptides of this list were detected using QSTAR/IDA – of them 148 (79.5%) were found in at least 2 of the 4 QSTAR runs (a percentage already lower than the QTRAP reproducibility across the four runs), 104 (55.9%) were found in at least 3 runs and only 63 (33.8%) (only 28.3% if we consider an universe of 223) of the peptides were found in all the runs – (Figure V.5B).

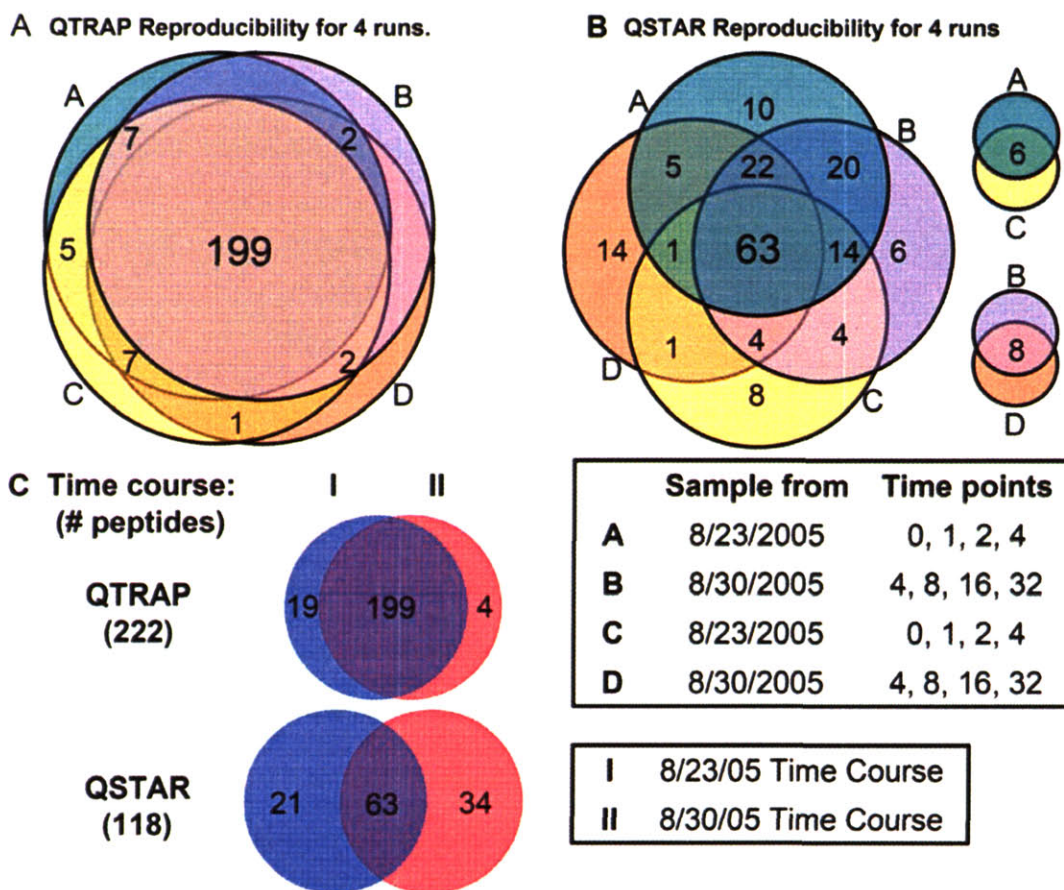


Figure V.5 – Reproducibility. A comparison of run to run reproducibility for 4 QTRAP/MRM and 4 QSTAR/IDA mass spectrometry runs shows that MRM is far superior in this respect. **(A)** Shows that of a total of 223 peptides (out of a list of 226 targets) 199 (89.2%) were found in all of the MRM 4 runs. **(B)** In contrast shows that of a total of 186 peptides (out of the same 226 list) only 63 (33.8%) were found in all of the IDA runs. **(C)** Shows that out of the 223 peptides found using MRM 222 (99.6%) were found such as a complete time course could be built for at least one of the two biological replicates analyzed, while out of the 186 peptides found using IDA only 118 (63.4%) were found such as a complete time course could be built for at least one of the two biological replicates analyzed. This number further reduces to only 52.3% when compared to the 223 peptides found by MRM. The high reproducibility of MRM allow for reliable quantification of phosphorylation and the accumulation of statistical significant data for mathematical modeling.

As mentioned before the four QTRAP and QSTAR runs are both biological replicates of a seven time point time course of HMEC cells following EGF stimulation. The data shows that of the 223 peptides detected using the QTRAP 222 (99.6%) were detected in at least two complementary runs (same sample,

different time points) allowing the construction of a peptide time course. Whereas of the 186 peptides detected using the QSTAR only 118 (63.4%) were detected in at least two complementary runs, only 53.1% of the total number of time courses acquired by adding the MRM method to our technology (Figure V.5C).

Assuming that all peptides are independent and equally likely to be found in any given run the probability of finding a desired peptide in any particular run is described by a binomial distribution. Although these conditions describe an ideal case and do not hold empirically they allow us to calculate the lower limit of reproducibility for these experiments. From our data the probability for successfully finding a desired peptide P is 0.9370 in the QTRAP/MRM case (95% confidence intervals [0.9193, 0.9518]) and 0.5446 in the QSTAR/IDA case (95% confidence interval [0.5117, 0.5771]). This means that for our universe of $N = 223$ peptides we expect to find 209 using QTRAP/MRM and 121 using QSTAR/IDA. If we now assume independency among runs (a much stronger assumption) the expectation of finding a peptide in 4 runs is 77.1% in the QTRAP/MRM and only 8.8% in the QSTAR/IDA. Clearly when it comes to reproducibility the QTRAP/MRM is expected to perform much better than the QSTAR/IDA. The fact that the expected values are lower than their empirical counterpart (77% v/s 89% QTRAP and 9% v/s 38% QSTAR) highlights the flaw in assuming that all peptides have the same chance to be found. In fact in both experiments more abundant peptides are more likely to be found, but this is really accentuated in the QSTAR because of the nature of IDA methods.

Intuitively speaking, by expanding our technology with the use of MRM we have developed a technique that is analogous to having an extremely, specific, quantitative, reproducible and high throughput (up to 960 quantifications per experiment) two dimensional western blot, were only identification with two exceptionally specific and independent antibodies at the same time (time resolution is the second dimension) is considered to yield a positive result.

V.3.5 Why 7 time points?

When trying to build predictive signaling models is important to cover as many conditions and have as good temporal resolution as possible. In chapter II we have shown that EGFR pY1148, ERK1 pY204, c-Cbl pY552 and MARVEL D2 pY23 belong to a group of phosphorylation sites that is activated immediately after EGF stimulation peaking after 5 minutes and decreasing monotonically until our last measurement at 30 minutes (also in (Zhang et al., 2005c)).

To normalize our present data at 5 minutes in order to be able to compare with our previous work on EGFR signaling, we have used MATLAB to fit a cubic polynomial to our 7 time points data. Figure V.6 shows in red our previous results and in blue our current time courses. It is clear from the figure that the activation of this 4 phosphorylation sites happens much more rapidly than our previous data showed. All of the sites seem to rise rapidly when stimulated reaching at least 70% of activation after only 2 minutes. In fact c-Cbl pY552 and MARVEL D2

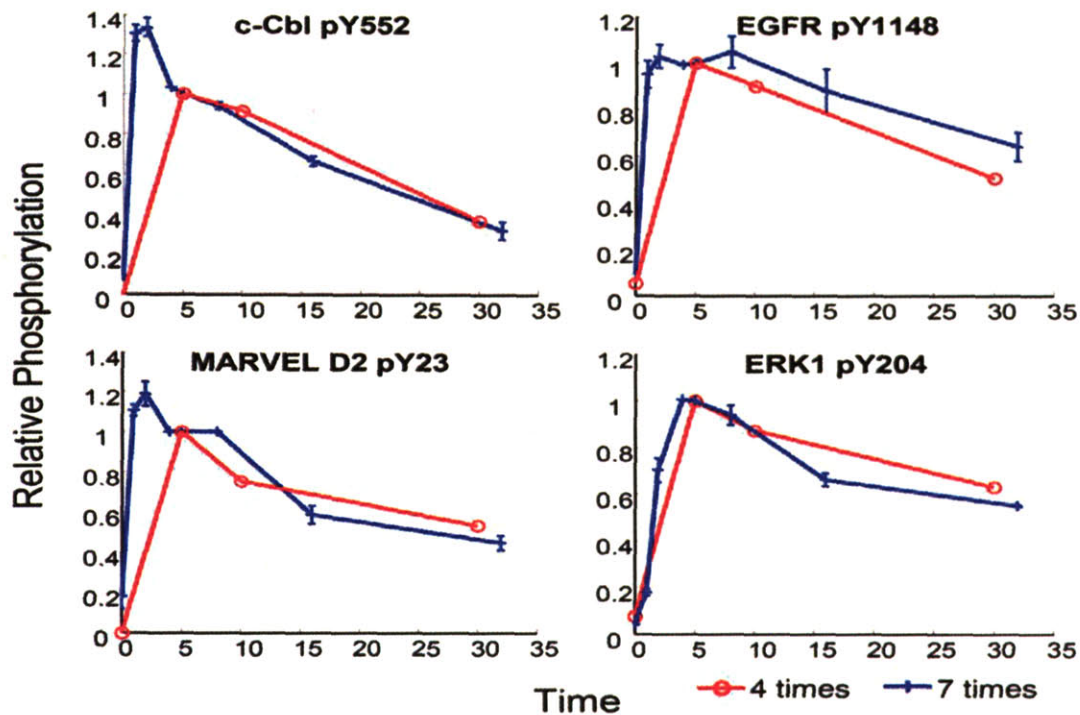


Figure V.6 – The Need for time resolution. This figure shows the temporal profiles of four phosphorylation sites (Cbl pY552, EGFR pY1148, Marvel D2 pY23 and ERK1 pY204) as quantified using our current 7 time points data set (blue) and a previous data set (Zhang et al., 2005c) with only 4 time points (red). It is clear from the profiles that increasing temporal resolution at early time points permits to distinguish dynamic traits of phosphorylation previously hidden to us (see Cbl pY552 and Marvel D2 pY23 profiles). Also the additional assessment of phosphorylation at 16 minutes allows for a better resolution of the de-phosphorylation kinetics of these four sites). This higher temporal resolution may allow in future works to build more accurate models of EGFR mediated phosphorylation dynamics.

pY23 reach maximal activation just 1 or 2 minutes after EGF stimulation, while EGFR pY1148 seems to reach a plateau between 2 and 8 minutes and ERK1 pY204 reach its maximum activity between 4 and 5 minutes.

In this example higher temporal resolution has allowed us to separate 4 phosphorylation sites related to early EGFR activation according to their dynamics, in fact just based on temporal dynamics we have been able to

separate c-Cbl and MARVEL D2 two proteins involved in degradation and internalization of EGFR (Grovdal et al., 2004; Sanchez-Pulido et al., 2002) from the receptor itself and ERK1. Even more c-Cbl and MARVEL D2 and EGFR and ERK1 can now be distinguished from each other base on their dephosphorylation after the 8 minutes time point, arguing for the presence of different mechanisms of dephosphorylation and/or degradation for these proteins. We expect that the high temporal resolution data we have generated here will be useful in future modeling efforts of the EGFR signaling network.

V.4 DISCUSSION

Although our new CSD technology once designed is very reliable and straight forward to use, three technical issues, two regarding instruments differences and the third regarding time profiling of peptides LC elution to adjust for MRM product ion detection limitations needed to be solved.

V.4.1 Resolution and Sensitivity

The first issue of concern when applying CSDMP is the different resolution/sensitivity between quadrupole - time of flight and triple quadrupole mass spectrometers. (While the first one performs at 10 fold higher resolution, allowing discrimination of peaks at less than 0.1 a.m.u distances triple quadrupoles allow for detection of 10 fold less material in a sample, detection being possible even in the low attomole range.) Because of their higher resolution QqTOF instruments allow for easy discrimination between iTRAQ labels and their contaminant shoulders. This information not available using triple quad instruments, must be taken from an ensemble of all correctly identified and quantified IDA hits for a given peptide precursor ion and used as a correction factor for MRM quantification when using the QTRAP. On the other hand because QTRAP instruments are at least 10 fold more sensitive, thus have a wider dynamic range than QSTAR and acquisition times for product ions in MRM methods can be tuned individually, peptides that were poorly quantified using

QSTAR instruments can be readily quantified using QTRAP instruments, solely by its intrinsic higher sensitivity and wider dynamic range or if needed, by increasing the accumulation time of product ions of scarce parent ions. It is these instruments intrinsic differences which generate the largest discrepancies between QTRAP and QSTAR quantification, but by taking advantage of both instrument strength highly accurate quantification can be obtained. In general, QTRAP quantification of peptides in the very low or high end of the QSTAR dynamic range (about 50 – 1500 iTRAQ counts in MS/MS) yields more accurate quantification, whereas samples with iTRAQ shoulder contamination are easier to quantify in the QSTAR.

A restriction to the manipulation of acquisition time of product ions in MRM mode is the desired duration of MRM cycles. Given that peptides might elute from an LC system into the mass spectrometer for only 30 seconds or less to even tens of minutes it is also necessary to take into account the elution time of all the peptides in the MRM method before defining its cycle time. As a rule of thumb the cycle time should be at least smaller than the width of the extracted ion chromatogram (XIC) of all the ions in the method and ideally smaller than one half or one third of the narrowest XIC to allow the acquisition of analytical replicates. Although in this paper, we set accumulation times for all product ions – b and y ions and iTRAQ labels – to 35 ms, amounting to a total cycle time of 12 seconds, in future studies using the same peptides it will be possible to modify the method time accumulation parameters to spend less time in the very abundant peptides and obtain even easier identification and quantification

signature of the scarce ones (by increasing their accumulation time), without changing the total cycle time more than a second or two.

V.4.2 Parent ion isolation and charge states

The second instrument related issue in expanding our technology by MRM was parent ion isolation. Preliminary runs comparing parent ion ionization in the QTRAP v/s the QSTAR showed that while in the second the most abundant parent ion for a peptide flies with charge between 3 and 5, in the QTRAP the most abundant parent ion present a charge usually lower than that shown for the same peptide in the QSTAR, in fact most peptides highest abundant parent ions presented charge 2 or 3 in the QTRAP. This charge differential was addressed by setting the MRM method parent ions in all the required cases to their m/z values at charge 2 or 3 with the restriction that no parent ion would be searched at m/z higher than 1200.

V.4.3 LC elution and MRM mode limitations

The final concern in expanding our technology towards monitoring mode is the QTRAP MRM method product ion detection limit which amount to a maximum of 300 ions per method. Although for traditional uses of MRM such as

drug detection in bodily fluids and absolute quantification of some of them or even in small scale proteomics experiments this is more than enough, iTRAQ based quantitative phosphoproteomics requires the identification of 6 product ions per peptide, setting the upper limit of using one MRM method to the identification and quantification of 50 peptides. In our study were more than 200 peptides (over 1200 product ions) were identified and quantified the use of 5 MRM methods in series was required for each analyzed sample. Of utmost importance to the use of more than one MRM method is to have as good knowledge as possible of all the peptides of interest relative LC elution time with respect to each other. In this way each method was programmed to search for peptides in order of elution, thus the first method looked for the early eluting peptides, while the fifth method looked for the late eluting ones. As an extra caution, and because elution order might vary from run to run every method started with the last 3 peptides (18 product ions) of the previous method. This overlap allowed us to move on from method to method quickly without fear of losing information and it proved to be a good feature as some peptides that were not found at the end of a method were usually found at the beginning of the next one.

Conceptually there is one more disadvantage when moving from IDA to MRM. While MRM is highly reproducible and peptide identification is highly accurate it still relies in the identification of 2 peptides fragments, one characteristic of the peptide N-terminus and the second characteristic of the peptide C terminus, in contrast IDA peptide identification in MS/MS, gives for

each peptide a comprehensive or at least a quasi-comprehensive list of the peptide fragmentation pattern, leading to sequence and reconstruction and PTM pinpointing. Notwithstanding this, if large enough b and y ions are chosen for MRM, their characteristic masses, coupled to parent ion information and the co-elution condition ensures that the right peptide has been identified. The difference is that in IDA we have a direct identification and in MRM identification is indirect. If this was really a concern the MRM method could be complemented with a confirmation experiment using the QTRAP third quadrupole as a linear trap for MS/MS identification of all the parent ions after MRM using enhanced product ion (EPI) mode for identity confirmation (Hopfgartner et al., 2004). The problem with using MS-MRM-EPI instead of just MS-MRM would be a large increase cycle time which would lead to a drop in the number of peptides detected and the number of analytical replicates for any peptide in a given run.

In this work we have shown the reliability of our expanded CSDMP technology. We have kept our highly accurate identification and quantification standards, while increasing our run to run selectivity and reproducibility from 33.8% to 89.2% across 4 experiments and our time resolution from 4 time points within 30 minutes to 7 time points within 32 – 4 of them in the time course first 4 minutes for early signal resolution.

By following IDA experiments with large scale MRM studies based on the IDA gathered information we were able to reproducibly identify and quantify a large section of the EGFR activated pathway including novel and reported phosphorylation sites with high temporal resolution. In fact by coupling in one

workflow IDA experiments and MRM experiments we have kept the advantages of discovery mode mass spectrometry, while adding the reproducibility benefits of monitoring mode mass spectrometry.

By combining IDA and MRM mass spectrometry we have developed a highly robust and quantitative technology for phosphoproteomics allowing the detection and monitoring of both novel and previously known phosphorylation sites in cell signaling networks. The robustness of our technology makes it ideal for the acquisition of large, consistent and quantitative, statistically significant protein phosphorylation datasets across many biological and dynamic conditions apt for rigorous computational analysis and mathematical modeling. Furthermore, the application of our technology to the study of clinical biopsies would allow the consistent monitoring of hundreds of signaling nodes in tumor samples, greatly improving our understanding of the protein biochemistry underlying cancer biology and facilitating recognition of dysregulated pathways and potential protein therapeutic targets in a patient to patient basis.

V.5 REFERENCES

Band V, Sager R (1989) Distinctive traits of normal and tumor-derived human mammary epithelial cells expressed in a medium that supports long-term growth of both cell types. *Proc Natl Acad Sci U S A* **86**: 1249-1253.

Grovdal LM, Stang E, Sorkin A, Madshus IH (2004) Direct interaction of Cbl with pTyr 1045 of the EGF receptor (EGFR) is required to sort the EGFR to lysosomes for degradation. *Exp Cell Res* **300**: 388-395.

Hopfgartner G, Varesio E, Tschappat V, Grivet C, Bourgogne E, Leuthold LA (2004) Triple quadrupole linear ion trap mass spectrometer for the analysis of small molecules and macromolecules. *J Mass Spectrom* **39**: 845-855.

Kim JE, White FM (2006) Quantitative analysis of phosphotyrosine signaling networks triggered by CD3 and CD28 costimulation in Jurkat cells. *J Immunol* **176**: 2833-2843.

Sanchez-Pulido L, Martin-Belmonte F, Valencia A, Alonso MA (2002) MARVEL: a conserved domain involved in membrane apposition events. *Trends Biochem Sci* **27**: 599-601.

Schmelzle K, Kane S, Gridley S, Lienhard GE, White FM (2006) Temporal dynamics of tyrosine phosphorylation in insulin signaling. *Diabetes* **55**: 2171-2179.

Stampfer MR, Bartley JC (1985) Induction of transformation and continuous cell lines from normal human mammary epithelial cells after exposure to benzo[a]pyrene. *Proc Natl Acad Sci U S A* **82**: 2394-2398.

Zhang Y, Wolf-Yadlin A, Ross PL, Pappin DJ, Rush J, Lauffenburger DA, White FM (2005c) Time-resolved Mass Spectrometry of Tyrosine Phosphorylation Sites in the Epidermal Growth Factor Receptor Signaling Network Reveals Dynamic Modules. *Mol Cell Proteomics* **4**: 1240-1250.

CONCLUSIONS

The work demonstrated on this thesis can be divided in two sections. Chapter I deals with a proof of principle study for anti-EGFR based antibodies therapeutics and chapters II, III and IV deal with development, application and refinement of mass spectrometry based technologies for phosphoproteomics.

Chapter I reveals that it is possible to downregulate EGFR activation using antibodies that block receptor dimerization without necessarily affecting receptor ligand binding. Production of such antibodies in a humanized form holds great potential for cancer therapy, especially for autocrine driven tumors where traditional ligand-competitive antibodies present low receptor inhibition efficiency.

Chapter II presents the development of a mass spectrometry based technology for quantitative phosphoproteomics. The new technology permits unambiguous identification and quantification of novel tyrosine phosphorylation sites in any given signaling cascade. In this chapter we applied the technology to detect and quantify the temporal dynamics of the cell signaling downstream of EGFR activation and we were able to use computational analysis to gain new biological insights on the network. Furthermore, the low amount of samples required make the methodology easily expandable to the study of cell cultures and in vivo biological samples including clinical biopsies. The study of biopsies

with this technology can have a profound impact on our understanding of cell signaling in cancer.

In Chapter III we used our newly developed technology to investigate the effect of HER2 in over-expression on signaling networks and corresponding cell behavior under EGFR and HER3 stimulation conditions. To correlate signals with cell response, proliferation and migration rates were also quantified for the same set of conditions. In human mammary epithelial cells HER2 over-expression promotes increased cell migration, but proliferation is mainly activated by EGFR not withstanding HER2 cellular levels. Computational analysis of the data revealed clusters of phosphorylation sites that are highly responsive to particular sets of cellular stimulation conditions, linking specific phosphorylation sites with activated receptor homo- or hetero-dimers. PLSR analysis of the phosphoproteomic and phenotypic datasets provided quantitative weighting to describe the relationship between specific phosphorylation sites and downstream biological behavior. As expected EGFR pY1173 strongly correlated with proliferation, while HER2 pY1248 strongly correlated with migration, but the analysis generated unexpected relationships as for example the correlation of KIAA1217 pY393 with migration or Dsc3a pY818 with proliferation. Overall the results of this chapter show – in addition to the new biological insights - the utility of our mass spectrometry technology for the analysis of multiple samples and the amenability of the data generated to mathematical modeling and computational

analysis. However the results also highlight the lack of run to run reproducibility of information dependent acquisition mass spectrometry.

In chapter IV we attempt to solve the low run to run reproducibility in our mass spectrometry technology by expanding it to use both information dependent acquisition (IDA) – which allows to work in discovery mode – and multiple reaction monitoring (MRM) – which allows to work in monitoring mode and is expected to enhance the method reproducibility –. We describe how to build an MRM method based on IDA results and we proceeded to validate the method by comparing its quantification results to the same results obtained by IDA. Adding MRM as a final step for our technology increases the reproducibility of peptides identified among four mass spectrometry runs from 33.8% to 89.2%. We therefore have improved our technology to be not only site specific and exquisitely quantitative, but also very reproducible and robust across different temporal profiles and biological samples. We applied this technology to a high resolution dynamic study of signaling downstream of EGFR activation and showed the importance of temporal resolution to distinguish between the phosphorylation profiles of many tyrosine sites in the network. Future applications of this technology hold great potential to acquire data amenable to computational analysis and mathematical modeling. The application of this technology to the study of clinical biopsies would allow the consistent monitoring of hundreds of signaling nodes in tumor samples, greatly improving our understanding of the protein biochemistry underlying cancer biology and

facilitating recognition of dysregulated pathways and potential protein therapeutic targets in a patient to patient basis.

Overall the work presented in this thesis has produced a couple of interesting new potential cancer therapeutic agents and a new site specific, quantitative and robust technology for the study of tyrosine phosphorylation mediated cell signaling networks. While developing this technique new biological insights had been uncovered on EGFR family activated networks. Finally, the potential has been built for detailed quantitative and robust studies of cell signaling with immediate applications to cancer biology and personalized therapeutics.

ACKNOWLEDGEMENTS

I want to start by thanking my funding sources (in chronological order) the DuPont Fellowship, the Biotechnology Processes and Engineering Center at MIT, the National Institutes of Health (NIH), the Biological Engineering Division at MIT and the Ludwig Graduate Fellowship for their support during the completion of my PhD.

I wish to express my thanks to all the people that made this work possible and so much fun to do. I want to thank MIT and the Biological Engineering Division for giving me this opportunity, I couldn't have asked for a better place or environment to study and work. In particular I want to thank JoAnn Sorrento for great conversations and chocolates, being extremely nice and helping me out with all my bureaucracy over these years. I have collaborated with several people through this thesis. The work presented in chapter II was done in collaboration with Ginger Chao and Mark Olsen. The work presented in chapter III was done in collaboration with Yi Zhang and the work presented in chapter IV was done in collaboration with Neil Kumar, Yi Zhang and Sampsa Hautaniemi. They had all been great collaborators and contributed extensively to this work, but in particular I want to thank Neil who was my closest collaborator, but most importantly an outstanding friend. I have not, but words of gratitude for all the members past and present of the White and Lauffenburger labs. Bart Hendriks, Kevin Janes, Lisa Joslin and Dan Kamei were instrumental in getting me started in the Lauffenburger lab, my thanks to them. Thanks also to Yi Zhang, Katrin Schmelzle and Paul Huang in the White lab for getting me mass spec able and ready to go.

I also wish to acknowledge my advisors Doug Lauffenburger and Forest White and my thesis committee (Linda Griffith and Dane Wittrup) for all their support and good advice. Doug has not only been a great scientific advisor, but he has also been a role model in many ways and a person I knew I could turn to for any kind of help or advice, for that and much more I want to thank him. About, Forest? He kind of got stuck with me, one day we are collaborating, the next day I am his first grad student – he did not have a choice! To be honest I couldn't have found a better advisor than Forest, his enthusiasm for science and every project he is involved in plus his constant laughter made working with him on a daily basis just awesome. He is been there to talk about science or anything else always with a positive attitude and always ready to help with whatever I needed. Through these years Forest has been an advisor a mentor and a role model, but even more important a great friend. I am sure this will continue even after I leave the lab, although I am still going to miss being there on a day to day basis.

Starting a PhD far from home in a foreign language is very tough. I really want to thank the dungeon crowd in particular Thomas Gervais (honorary), Nate Tedford and Maxine Jonas for pulling me through a not so easy first year, being

such good friends make me laugh so hard and teaching me their version of english (Nate). I also got the best desk, office and lab, so thanks to Lisa Joslin, Melissa Kemp, Nicholas Marcantonio, Neil Kumar, H-D Kim, Maria Ufret, Paul Huang, Leo Iwai, Amy Nichols, Hideshiro Saito-Benz and Katrin Schmelzle for all the science and fun. I also want and need to thank my roommates and my roommate in law Caitlin Bever – after all this years my foster sister –, Heather Keller and Ty Thomson; they had been just great to live with, they had been like a home away from home for me. They have “never” made fun of, or laugh at me and I know they are friends I’ll be able to count on all my life (sorry guys I’ll keep bothering you for a long, long time!). My many, many (and I mean it) soccer teams had been able to keep my feet on the ground – sorry, ball , but in particular I need to mention my favorite player, great friend and evil tackling machine Danielle France – nothing better when I am down or happy than megging Danielle, thanks D!. I have made many friends here and I want to thank you all. I could have never imagined how much fun doing a PhD could be and how many cool people would I get to meet. I can’t mention them all here but you are definitively included in this list... yes YOU! Well may be not you, but you definitively yes!

Finally, I need to thank my family. They are great. I have been lucky to have an incredible sister Andrea Wolf Yadlin and an awesome brother Rodrigo Wolf Yadlin, they have always supported me and push me forward when I was down or tired, they make me laugh, they keep me honest and they always show how much they love me and they make a point of making me aware of how lucky am I to have them! Thanks for all, I love you guys. My grandfather Bernardo Yadlin and my grandmother Chaque Abeliuk just come from another galaxy; they have taught me so much and made me so happy that it is hard to express, they are the best people ever and I love them. My mom Yudith Yadlin is a force of nature of her own. She is far too good. She took care of my siblings and me when my dad died and she is been selfless all her life. She has always believed in me and she has stood behind me and pushed me forward. Her support and love had taken me where I am today and I have nothing but admiration and love for her.

DEDICATIONS

I want to dedicate this thesis to my mother Yudith Yadlin, my grandfather Bernardo Yadlin, who passed away while I was completing my PhD and his wife, my grandmother Chaque Abeliuk. There are no words that I can use to describe what their mean to me. They have during all my life been an example to emulate and they have taught me how to act and how to be not by words, but by example. They are honest, intelligent and humble persons, always giving never taking and always worried about people over anything else. They have taught me to work hard, to enjoy life, to care about people, to always be honest to others and to myself, to be happy, to laugh, to love and to make the best of everything and everyone. They are the corner stone of my family and wherever I go I think of them. I feel it is well suited to dedicate this thesis to them because they have always loved and supported me. They have made me proud and happy to be their son and grandson since I can recall and I just aspire to one day be like them.

Gracias,
Los quiero mucho,
Ale

INDEX OF TABLES AND FIGURES

I. INTRODUCTION

Figure I.1.	Crystal structure of EGFR-EGF homodimer	19
Figure I.2.	Crystal structure of auto-inhibited EGFR bound to EGF	19
Figure I.3.	Crystal structure of HER2	20
Figure I.4.	The ErbB signaling network	23
Figure I.5.	Schematic of a QSTAR instrument	49
Figure I.6.	b and y ions	52
Figure I.7.	Schematic of a QTRAP instrument	54
Figure I.8.	SRM/MRM method schematic	55
Figure I.9.	Resolution	57
Figure I.10.	iTRAQ chemistry	66

II. ANTIBODIES AGAINST DOMAINS II AND IV OF EGFR INHIBIT RECEPTOR PHOSPHORYLATION AND DECREASE HIGH-AFFINITY BINDING BY BLOCKING PREFORMED DIMERS

Figure II.1.	EGFR epitopes selected for peptides mimics and antibody harvesting	110
Figure II.2.	Specificity of rabbit polyclonal antibodies	111
Figure II.3.	Effect of Anti-DII and Anti-dIV polyclonal antibodies in EGFR activity measured as pY1173 auto-phosphorylation	112
Figure II.4.	The polyclonal antibodies block homo- and hetero-dimerization	113
Figure II.5.	Binding of Anti-DII and Anti-DIV antibodies does not affect EGF binding to EGFR on HMECS	114
Figure II.6.	Equilibrium titration and Scatchard plots show that Anti-DII and Anti-DIV antibodies abrogate the high affinity binding component of EGF-EGFR	115

III TIME-RESOLVED MASS SPECTROMETRY OF TYROSINE PHOSPHORYLATION SITES IN THE EGF RECEPTOR SIGNALING NETWORK REVEALS DYNAMIC MODULES

Figure III.1.	Mass spectrometry-based approach to quantitative analysis of phosphorylation dynamics on specific tyrosine residues	135
Figure III.2.	Representative MS/MS spectra for two phosphorylated peptides from proteins not previously reported in EGFR signaling	139
Figure III.3.	Comparison of tyrosine phosphorylation dynamics as measured by mass spectrometry and western blot analysis	142
Figure III.4.	Self-organizing maps (SOMs) facilitate the identification of EGFR signaling modules	147
Table III.1.	Quantitative Time Course Tyrosine Phosphorylation Profiles	136
Table III.2.	Tyrosine-phosphorylated proteins not previously reported in EGFR signaling	138
Table III.3.	Tyrosine Phosphorylation Sites not previously reported in the literature	140

IV QUANTITATIVE PHOSPHOPROTEOMIC ANALYSIS OF HER2-OVEREXPRESSION EFFECTS ON CELL SIGNALING NETWORKS GOVERNING PROLIFERATION AND MIGRATION

Figure IV.1.	Data acquisition and quantification scheme with example data	178
Figure IV.2.	Hierarchical Clustering of phosphorylated peptides to identify co-regulated phosphorylation sites.	181
Figure IV.3.	Self Organizing Map to cluster phosphorylation sites across conditional and temporal space	186
Figure IV.4.	Self Organizing Map (SOM) Component planes	187
Figure IV.5.	EGF and HRG drive migration and proliferation to varying extents in HMEC parental and HER2 over-expressing cells.	191

Figure IV.6.	Effect of increased HER2 expression on phosphorylation sites within the EGFR signaling network	193
Figure IV.7	Effect of increased HER2 expression on phosphorylation sites within the EGFR signaling network 2	194
Figure IV.8.	Effect of HRG stimulation versus EGF stimulation in high HER2 expressing cells	199
Figure IV.9	Effect of HRG stimulation versus EGF stimulation in high HER2 expressing cells 2	201
Figure IV.10	Partial least squares regression correlates 248 protein metrics to cell migration and proliferation	203

V MULTIPLE REACTION MONITORING AND INFORMATION DEPENDENT ACQUISITION: TOWARDS A REPRODUCIBLE, QUANTITATIVE AND ROBUST MASS SPECTROMETRY BASED TECHNOLOGY FOR PROTEOMICS

Figure V.1.	Schematics of CSDMP Method	232
Figure V.2.	How to construct MRM method from IDA data	234
Figure V.3.	Detecting and quantifying protein phosphorylation by MRM	238
Figure V.4.	Validation of MRM quantification	241
Figure V.5.	Reproducibility	244
Figure V.6.	The Need for time resolution	247

ABBREVIATIONS

HMEC:	Human Mammary Epithelial Cell
EGF:	Epidermal Growth Factor
HRG:	Heregulin
EGFR:	Epidermal Growth Factor Receptor
HER2:	Human Epidermal Growth Factor Receptor 2
HER3:	Human Epidermal Growth Factor Receptor 3
LC/MS/MS:	Liquid Chromatography tandem Mass Spectrometry
pY:	Phosphotyrosine
IMAC:	Immobilized Metal Affinity Chromatography
QqTOF-TOF:	Quadrupole – Time of Flight
QSTAR:	A particular QqTOF-TOF instrument
Triple Quad:	Triple Quadrupole
QTRAP:	A particular Triple Quad instrument
IDA:	Information Dependent Acquisition
MRM:	Multiple Reaction Monitoring
CE:	Collision Energy
CAD:	Collision Activated Dissociation

PERMISSIONS

Figure I.4 was Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Molecular Cell Biology 2: 127-137, copyright 2001.

NATURE PUBLISHING GROUP LICENSE TERMS AND CONDITIONS

Aug 10, 2006

This is a License Agreement between Alejandro M Wolf Yadlin ("You") and Nature Publishing Group ("Nature Publishing Group"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

License Number	1525370136877
License date	Aug 10, 2006
Licensed content publisher	Nature Publishing Group
Licensed content publication	Nature Reviews Molecular Cell Biology
Licensed content title	UNTANGLING THE ErbB SIGNALLING NETWORK
Licensed content author	Yosef Yarden, Mark X. Sliwkowski
Volume number	2
Issue number	2
Pages	127-137
Year of publication	2001
Portion used	Figures
Number of figures	1
Requestor type	Student
Type of Use	Thesis / Dissertation
Total	\$0.00

Figure 1.10 was Reprinted by permission from ASBMB Journals: Molecular Cellular Proteomics 3: 1154-1169, copyright 2004

ASBMB Journals

Journal of Biological Chemistry

Molecular and Cellular Proteomics

Journal of Lipid Research

Biochemistry and Molecular Biology Education

ASBMB Today

ASBMB does not charge for and grants use without requiring your copyright permission request for:

- Original authors wanting to reproduce portions of their own work; or to republish their material in not-for-profit formats or venues.
- Students wanting to reproduce or republish their work for educational purposes.
- Students using other authors' material for their theses.
- Reproduction or republication of abstracts only.
- Photocopying up to 5 copies for personal use.
- Non-profit educational institutions making multiple photocopies of articles for classroom use; all such reproduction must utilize institutionally owned equipment for this purpose.

Use of copyrighted material requires proper citation.

http://www.mcponline.org/misc/Copyright_Permission.shtml