

# Chapter 2

## Modeling Signaling Networks Using High-throughput Phospho-proteomics

Camille Terfve and Julio Saez-Rodriguez

**Abstract** Cellular communication and information processing is performed by complex, dynamic, and context specific signaling networks. Mathematical modeling is a very useful tool to make sense of this complexity. Building a model relies on two main ingredients: data and an adequate model formalism. In the case of signaling networks, we build mainly upon data at the proteome level, in particular about the phosphorylation of proteins. In this chapter we review recent developments in both data acquisition and computational analysis. We describe two approaches, antibody based technologies and mass spectrometry (MS), along with their main features and limitations. We then go on to describe some model formalisms that have been applied to such high-throughput phospho-proteomics data sets. We consider a variety of formalisms from clustering and data mining approaches to differential equation-based mechanistic models, rule-based, and logic based models, and on through Bayesian network inference and linear regressions.

### 1 Introduction

Whatever their nature, identity, and environment, cells are continuously exposed to signals, whether reflecting their internal state, or emerging from growth factors, neighboring cells or the extracellular matrix. All these signals need to be received, interpreted and possibly transmitted or propagated, in an integrated manner so

---

C. Terfve  
European Bioinformatics Institute (EMBL–EBI), Wellcome Trust Genome Campus,  
Cambridge CB10 1SD, UK  
e-mail: [terfve@ebi.ac.uk](mailto:terfve@ebi.ac.uk)

J. Saez-Rodriguez (✉)  
EMBL–EBI and European Molecular Biology Laboratory (EMBL), Genome Biology Unit,  
Meyerohofstrasse 1, D-69117 Heidelberg, Germany  
e-mail: [saezrodriguez@ebi.ac.uk](mailto:saezrodriguez@ebi.ac.uk)

as to produce the appropriate response. This information processing is performed through the use of highly dynamic and context specific networks assembled from a multitude of signaling molecules [32]. Given their fundamental role in cellular function and intercellular coordination, deregulation of signaling networks is often involved in the development of diseases [3, 27]. Furthermore, although development of resistance to drugs can happen through accumulation of mutations, it seems that another underlying mechanism can be the rewiring and adaptation of the signaling network. Combinatorial genetic perturbations in yeast suggest that signaling networks are extremely adaptive to such perturbations [32]. Studying signaling networks as a whole, in physiological and disease contexts, is therefore essential to understand how cells function and respond to their environment and how this process is deregulated in diseases, to potentially provide new venues for therapies.

Understanding how the elements that make up signaling systems are organized and function together to allow the cell to respond to a perturbation is a challenge. This is only the beginning that is to be investigated [28]. Therefore, it has been argued that mathematical modeling is necessary to make sense of the sheer amount of elements that enter into play [1, 13, 28, 32, 40]. A key point when modeling is to be aware of the assumptions made in building the model (level of detail, scope, etc.), and to interpret the model outputs correctly [5, 65]. Indeed, there are many ways to model biochemical and in particular signaling networks, and the choice of a particular formalism depends on the system under investigation, the data at hand, and the question to be answered.

Ideally, one builds a model based on good quality data of the system under study; in the case of signal transduction data is at the proteome level, since proteins are the ultimate agents of cellular activity. Events occurring at the transcriptional levels can however also be included in models of signaling systems (see [34] for a review on cellular regulatory networks encompassing regulation at different levels). Proteomics is the field of biology that studies the expression, modification, conformation, and activity of proteins in a system [2, 30]. This is challenging for many reasons, mainly because screening of the entire proteome (as can be done for the genome or transcriptome) is impractical with current technology and because the proteome cannot be defined using a list of proteins. Indeed, the wide varieties of post-translational modifications (PTMs) and their combinations lead to a combinatorial explosion of the number of states that need to be assessed. To add to the complexity of this picture, protein abundance in a single cell population frequently spans more than 6 orders of magnitude [2], which is broader than the dynamic range of any routine proteomics technology currently available. In addition, limited information can be inferred from investigations at the mRNA level since transcripts levels are poorly correlated with protein abundance [23, 73]. An ideal technology to probe the proteome of human cells in a signaling framework would have to be able to measure accurately the concentrations of more than 30,000 different proteins and their splice variants, each possibly subject to a variety of post-translational modifications (e.g., an estimated 100,000 sites for phosphorylation alone) and should be able to measure all this in a time-dependent,

cell and compartment specific manner, under various conditions such as genomic background and stimulations by drugs [62]. Although no single platform is currently capable of such an achievement, some progress has been made with regards to capabilities of medium to high-throughput proteomics technologies, which we will discuss further in this chapter. In particular, we will discuss advances in antibody based and mass spectrometry (MS) based proteomics workflow that have allowed and will allow informative modeling of signaling networks.

Signal propagation involves changes at three levels: regulated PTMs, protein–protein interactions (often owing to PTMs), and changes in the expression level of proteins. These three levels are coordinated through dynamic regulation which is often spatially segregated [10]. Amongst these three levels, we focus on PTMs because they are the most immediate events, often trigger the other events and can thus often be used as a proxy for those other events. However, depending on the time scale of the process studied, it is possible that expression and degradation events would play an important role in the dynamics of the system. One should therefore be cautious and, if possible, also be able to measure the abundance of proteins.

Over 200 types of PTMs have been reported and that number is still growing [23]. In particular, phosphorylation and associated players (kinases, phosphatases, and phospho-binding domain containing proteins) play a very important role in signaling since this PTM can control the formation of multiprotein complexes, the dynamic localization of proteins, as well as their stability and enzymatic activity, and about 30% of proteins are phosphorylated at any one time [13]. This points to the importance of phosphorylation in the context of signaling, and combined with the availability of assays to measure phosphorylation of proteins on a large scale, this makes phospho-proteomics a common focus when looking at signaling.

The phosphorylation state of a protein reflects the result of the action of kinase/phosphatase reaction pairs. In some cases, e.g., phosphorylation of the activation loop of some protein kinases, this is correlated with the activity status of the protein. However, this is not true in the general case, and care should be taken when interpreting phosphorylation status [13]. Relatively recent data suggests that most *in vivo* phosphorylation sites have not even been detected yet [54]. In addition, although we can now sequence a full genome, signaling data are necessarily incomplete [20] because characterizing a “full” signaling network would necessitate to either characterize the full state of the cell (including contextual information and state of modification/interactivity of every agent) or to be able to delineate where the signaling network starts and ends.

PTMs not only have individual roles but can also function in combinations to precisely regulate molecular interactions, protein activity and stability, in a context specific manner [29]. Therefore, interpreting phosphorylation in a signaling context is likely to prove very challenging, and prioritizing functionally important phosphorylation sites for experimental investigation is going to be crucial [67], as will be the identification of kinases involved in particular modifications [13].

This chapter is outlined as follows: in the first part, we will discuss available experimental platforms and explain their particular features and limitations. In the second section, we will briefly discuss methods for building signaling networks that

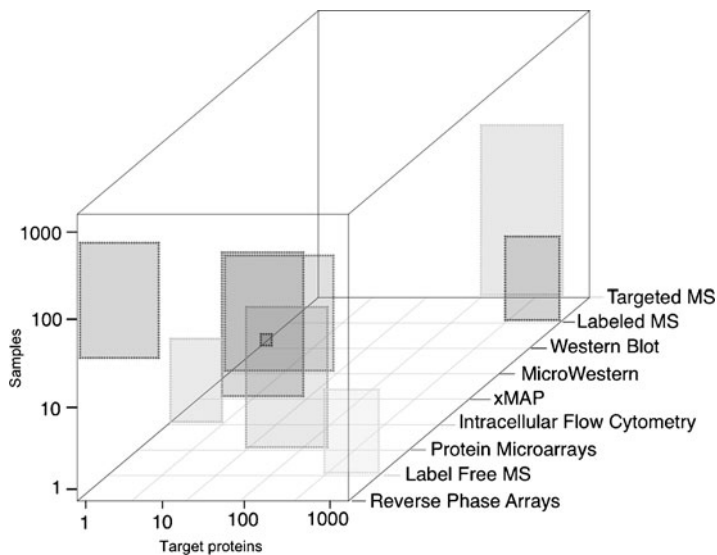
have been applied to high-throughput phospho-proteomics data sets. This section is not exhaustive neither in terms of application examples nor in terms of specific methodologies, but it rather aims at describing modeling frameworks that have been applied to phospho-proteomics data by showing a couple of examples, and discuss the advantages and disadvantages of each method. Most of the methods presented have been applied in many other contexts (often prior to signaling networks), in particular to the much more mature field of modeling of gene regulatory networks [4, 31, 44].

## 2 Phospho-proteomics Data Collection

Data collection methods for medium-/high-throughput proteomics can be roughly divided into two categories: those that do not make any assumption about the sample composition (e.g., shotgun MS), and those that measure a predetermined set of proteins (e.g., affinity based approaches) [2, 62]. Affinity based technologies most commonly make use of antibodies, and those methods will be the subject of the first part of this section. The second part of this section will examine the principles of common shotgun MS and will take a closer look at targeted MS as a potential alternative to antibody based approaches to generate large data sets for the development of systems biology models. The choice of a method to use ultimately depends on the material and expertise available, and the number of experiments that can be performed often results from a balance between the time and cost per experiment.

### 2.1 *Antibody-based Methods*

All antibody based methods build upon the same principle: the interaction of a target protein with an antibody, an interaction that should happen with both high-affinity and selectivity. Therefore, all of these methods suffer from the same limitation: the data is only as good as the antibodies are, and investigators are therefore limited by the availability of high-quality antibodies [2]. However, new multiplexing technologies offer the ability to analyze hundreds to thousands of samples a day, thereby allowing assays on multiple time points, and across multiple conditions of interest (which is not yet possible with MS due to the labor intensive process of analyzing more than a few conditions), although the total number of signals measured rarely exceeds a few dozens [3, 75] (see Fig. 2.1 for an overview of multiplexing capacities of the methods examined in this chapter). Such data sets usually need to be normalized and quality controlled (e.g., assessing reproducibility, detecting outliers, etc.). There are a number of computational tools to do so and to connect processed data to modeling tools. This is however outside the scope of this chapter.



**Fig. 2.1** Overview of multiplexing capacities of data collection methods. This figure displays approximations for the ranges of numbers of samples and proteins that can be interrogated using various phospho-proteomics platforms. These numbers are not intended as absolute statements on the capabilities of these methods and in particular they may change slightly depending on the actual protocol used, experimental setting and proteins targeted. This figure is inspired from [2]

The combination of fluorescently labeled antibody recognition and single cell measurement capacity of fluorescence activated cell sorting (FACS) [59] seems to be a promising technology due to its single cell nature. Another promising technology is the microwestern array developed by Ciaccio et al. [11], which in addition to antibody based recognition provides an extra separation step by electrophoresis. Other technologies, which we will not discuss in this chapter, such as high-throughput microscopy [24] associated with immunofluorescence, and mass cytometry [68], are being developed and could potentially be applied to the generation of high-throughput phospho-proteomics data sets adapted for modeling. To date, most commonly used antibody-based technologies are protein arrays, reverse-phase protein arrays, and the bead based xMAP technology from Luminex [62].

### 2.1.1 Intracellular Multicolor Flow Cytometry

Intracellular multicolor flow cytometry allows the simultaneous measurement of multiple phosphorylated proteins and phospholipids in large populations of cells, on a single cell basis [59]. The principle is simple, the cells are fixed and incubated with fluorescently labeled antibodies, and then are subjected to FACS which quantitatively measures the targets' expression or modification level. The main limitation of this technology is the availability of suitable reagents, i.e., antibodies compatible

with flow cytometry. Furthermore, this technology only allows a relatively small number of proteins to be examined simultaneously (up to a dozen). The ability to barcode cell populations before protein labeling potentially allows this technology to be applied to the processing of multiple samples/conditions in parallel [38].

### 2.1.2 Microwestern Arrays

Microwestern arrays build upon the well established western blot technology, which enables quantitative and sensitive analysis of protein abundance and modification after electrophoretic separation, while a high-throughput capacity is achieved by applying the protocol to microarrayed cell lysates. The main advantage of adding the electrophoretic separation step to the workflow is that it allows for a reduction in sample complexity, whereas the antibody detection step results in signal amplitude proportional to the abundance of immobilized protein. The signal localized at a physical position on the membrane can be related to molecular size standards, so the antibody cross reactivity problem associated with most other technologies can be controlled to some extent [11]. This method showed, in the proof of principle study, a linear relationship between antigen concentration and signal intensity over from 2 to 3 orders of magnitude [11]. The main advantages of this method over classical protein arrays are an increased specificity owing to the electrophoretic separation step, low sample requirements (compared to technologies such as xMAP) and the wide availability of reagents since antibodies developed for the classical western blot should be applicable to this method.

### 2.1.3 Array and Bead-based Methods

All other methods described here (reverse phase arrays, protein arrays, and xMAP technology) rely on the same principles and in particular are composed of three main ingredients: (1) an identification system which is required for multiplexing (i.e., a physical support with unique identity, whether a location on a 2D arrangement or unique physical properties of beads in suspension), (2) a capture system (to immobilize the protein(s) of interest, whether directly on the support as in reverse phase arrays or through interaction with antibodies as in protein arrays), and (3) a detection system (to produce a signal that is ideally linearly proportional to the amount of captured target protein, typically fluorescent-labeled detection or enzymatic-labeled detection such as a biotinylated secondary antibody bound by a streptavidin-linked peroxidase) [2].

In protein microarrays, the captured antibody is covalently bound to a slide in an ordered manner, and the slide is incubated with the sample. For detection, either the sample itself is chemically labeled with a fluorophore, or it is detected by a labeled secondary antibody (sandwich assay). This technology can measure up to hundreds of proteins but the number of samples is somewhat limited. Direct labeling allows for the simultaneous measurement of multiple analytes and only requires one

high-quality antibody per target protein, but due to uneven labeling of all proteins and chemical alterations this method can be rich in false positive and display a high-background. Sandwich assays on the other hand provides a more accurate and specific detection, but require two high-quality antibodies [75]. This is not a trivial problem for microarrays as, contrary to antibodies for western blots that detect denatured proteins, antibodies for such array technologies must be able to recognize the substrate in native state but immobilized on a slide, which can impose steric constraints on the interactions.

Reverse phase arrays are similar in principle but in this case the lysate itself is spotted on the support and therefore multiple lysates (dozens to hundreds) can be processed on a single slide. One can then either incubate the entire slide with one antibody or create physical compartments within which distinct primary antibodies can be used. A labeled secondary antibody then binds the captured antibody. This technology only requires one specific antibody for detection of each protein but it is therefore highly dependent on the selectivity of this antibody, and this added to the presence of all cellular proteins bound to a slide is bringing up issues of cross reactivity that have been reported to cause substantial noise [2,11,75]. Therefore, the accuracy of reverse phase arrays tends to be lower than that of protein microarrays, specifically when sandwich assays are used [58].

The xMAP technology is conceptually similar to protein microarrays except that rather than being localized on particular spots on a support, specific antibodies are associated with microspheres in suspension that are internally dyed to generate different spectral signatures. This technique theoretically supports the analysis of up to 100 analytes per well, since the beads can be multiplexed and incubated with a single sample. For detection, a mixture of biotinylated antibodies is added, and a fluorescently labeled molecule binds the detection antibody. Quantification is obtained by a flow cytometer based instrument capable of reading the beads' spectral signature and the fluorescence intensity simultaneously. Having beads in suspension rather than planar microarrays allow for faster reaction kinetics and high-surface to volume ratio, and consequently better washes and homogeneous chemical reactions resulting in an increase in the signal to noise ratio [2,71]. A disadvantage of this approach compared to protein arrays is that it requires considerably more cell material and the cost of detection is approximately 30 times higher per protein detected [11].

## 2.2 *Mass Spectrometry*

MS is an analytical technique that determines the mass to charge ratio of charged analytes, thereby providing a means to identify chemical compounds. Applied to proteomics, it allows systematic protein identification and quantification (provided that an appropriate protocol is used) from complex samples using a combination of liquid chromatography separation of peptides generated by digestion, followed by their analysis by tandem MS (a protocol called shotgun LC-MS/MS) [23].

### 2.2.1 Shotgun MS/MS

A classical shotgun MS workflow proceeds as follows: the protein samples are digested with trypsin and the lysate is fractionated by reversed-phase liquid chromatography, which is used to separate the complex mixture of peptides on the basis of their hydrophobicity. Other types of chromatography such as strong cation exchange are also commonly used, where the peptides are separated based on their charge. The peptides in fractions eluting from the chromatography columns are then vaporized and ionized, typically by subjecting the solution to an electric potential, which causes the formation of a spray and the desolvation and ionization of the peptides (a technique called electrospray ionization) [23]. In the MS stage, the mass to charge ratio of all ions is determined, then the first mass analyzer selects ions for collision induced dissociation, where neutral gas molecules are used to fragment the peptide. The resulting fragment ions are measured in the second mass analyzer of the tandem MS [10, 23]. The precursor ion intensities measured at the MS stage can be used for peptide quantification, and the MS/MS fragment ion information can be used to identify the peptide through its sequence, by comparing the experimental MS/MS fragmentation pattern to theoretical counterparts derived from a database of sequences from *in silico* digested proteins. Subsequent protein identification can be obtained through a database search [23]. For a review about how to obtain and interpret sequence information from tandem MS experiments, we refer the interested reader to reference [66].

### 2.2.2 Data Processing Challenges

The problem of assigning sequences to MS spectra is not a trivial one, and each identification should be carefully assessed for its statistical significance [8]. Most of the algorithms performing this task report one or more peptide spectrum match (PSM) scores that reflect the quality of the match between the experimental and computed theoretical peptide spectrum. Statistics associated with these scores are typically obtained by searching the data against a target/decoy database, i.e., in addition to search through real sequences, the search is also performed against a randomized, shuffled, or reversed database. This gives an approximation of the FDR (expected proportion of false assignments among a selected set of predictions) by counting the number of matches in the target (presumably mainly true positives), and decoy (presumably mainly false positives) databases that satisfy a score criteria. Some algorithms supplement this information by implementing methods to improve the discrimination between correct and incorrect PSMs, for example, by building classifiers that also make use of other features reported by the search algorithm, such as charge state, difference in score to the second best hit, etc., which are often used by experts to manually validate the PSMs [8].

After having identified the peptide present in the sample with a certain level of confidence, another problem arises before the data can be readily interpreted: the protein assignment problem, i.e., identifying the protein composition of the sample



from which the peptide sequences result [52]. Indeed, the same peptide sequence can match multiple different proteins, making both identification and quantification challenging. This problem can be partially alleviated when the sample complexity is further reduced prior to digestion and LC–MS/MS using techniques such as 2D gels, which can provide additional information such as molecular weight and the isoelectric point of the protein. The issue is particularly challenging in the case of higher eukaryote organisms since these organisms present a certain degree of sequence redundancy [52].

Distinguishing between different proteins of similar sequence is of course increasingly difficult when the sequence coverage decreases (i.e., the fraction of the protein sequence that is covered by identified peptides). Unfortunately, the sequence coverage observed in shotgun MS proteomics experiments is typically quite low. Several factors contribute to this, such as, the size of the proteins to be identified, enzymatic digestion constraints, and the detection mass range of the instrument. Furthermore, some unexpected PTMs can lower the chances of a peptide being observed, and low abundance or poorly ionizing peptides are also less likely to be selected for MS/MS sequencing [52]. For more information about this topic, we refer the reader to the following review [52].

### 2.2.3 Quantitative MS

Regarding the quantification of proteins using MS, two main approaches can be applied: differential isotope labeling and label free quantification. Differential isotope labeling builds on the hypothesis that when measuring two analytes of identical chemical composition but different stable isotope composition, their relative signal intensity represents their relative abundance in the sample. There are two main ways to do this: *in vitro* labeling or *in vivo* incorporation of isotope-labeled amino acids through metabolic labeling (stable isotope labeling with amino acids in cell culture, SILAC). For *in vitro* labeling, the two samples are prepared separately and the protein or peptide solutions are individually labeled with heavy or light version of tagging reagents. The recently introduced iTRAQ technology allows peptide labeling with isobaric tags, as the name indicates, keeps the mass of differentially labeled precursors constant, i.e., appearing as a single peak in the MS1 spectrum. Quantification occurs in the MS/MS spectrum by comparing peak areas of sample-specific reporter ions [23, 58]. Compared to isotope labeling techniques which only allow up to typically three samples to be compared simultaneously, the iTRAQ labeling protocol can compare up to eight samples in a single LC–MS/MS run. A very similar idea is implemented in the Tandem Mass Tags protocol [70]. SILAC is an *in vivo* labeling method where different populations of cells are grown in presence of media containing light or heavy isotope versions of lysine or arginine most commonly [23], although other amino acids have been used (e.g., leucine [55]), and labeling of living animals such as rats with N15 has also been reported [22]. Since the labeling occurs very early on in the protocol, this method avoids many

of the errors and biases than can be introduced in the sample processing. However, this method is limited to cells or organisms that can be metabolically labeled, i.e., typically cell cultures and not primary samples [23], and it is quite a complex and time consuming protocol, which limits its implementation to laboratories with significant infrastructure [13].

Label free quantification by peptide precursor ion intensities is based on the alignment of high-mass accuracy MS1 (i.e., precursor ions) spectra obtained from separate LC–MS/MS experiments. Peptides are identified and aligned based on their specific retention time and mass to charge ratio. The relative abundance changes are calculated from the aligned spectra on the basis of the signal intensities of extracted ion chromatograms. Another label free method, spectral count, relies on the assumption that the rate at which a precursor ion is selected for fragmentation is correlated to its abundance. The spectral counts from peptides mapping to the same protein are then averaged into a protein abundance index. This method depends on the quality of the MS/MS peptide identification and protein assignment, and although it works relatively well for abundant proteins, it is often problematic for small and low abundance proteins [23]. In general, label free techniques provide a less accurate quantification than stable isotope label methods [58].

## 2.2.4 MS for PTMs

Because the addition of a PTM to a protein causes a defined mass change, MS can measure and localize modifications with a single amino acid resolution. However, PTM analysis poses specific challenges beyond those described above: modified peptides are often present at low amounts, can lead to more complicated MS/MS spectra and increase the database search space [10]. Therefore, it is usually necessary to enrich the sample for the modification of interest in order to increase the dynamic range and sensitivity. Depending on the PTM, this can be done by derivatization of the PTMs and chemical solid phase capture, or more commonly, for phosphorylation using metal affinity chromatography, titanium dioxide chromatography, or antibodies specific for a modification [23]. Ideally, one would hope to obtain all modified peptides and only those but in practice all modified peptides will be enriched to a certain degree with respect to the starting mixture, with an enrichment factor that can range from only several folds for some modifications to over a hundred fold for phosphorylations [10].

When looking at PTMs, two different tasks are performed: the identification of the peptide bearing the PTM, and the unambiguous localization of the PTM-bearing amino acid on this peptide [10]. Neither of these tasks is trivial, and although in principle any PTM can be detected provided that it leads to a modification in the mass of the peptide, in practice a full mapping of the PTMs of a protein requires full sequence coverage (i.e., detection of all the peptides of the protein). This is not straightforward as typically only a subset of the peptides generated by proteolytic digestion of a protein are detected, unless optimization strategies are used.

### 2.2.5 Limitations of the Shotgun MS/MS Approach

Although shotgun MS/MS approaches offer a coverage of the proteome that no other technology can currently approach (i.e., about 7,000 proteins can be quantified in an experiment) [10], the technology also shows several limitations. A first limitation is that this depth of analysis typically comes at a high cost in terms of time of experiment (i.e., experimental time typically in days), which limits the ability to interrogate multiple conditions/samples. For this reason, classical shotgun proteomics workflows are better qualified as “high-content” than “high-throughput” experiments. Other fundamental limitations are extreme redundancy and under sampling associated with the method, which result in a saturation effect, i.e., the number of proteins currently identified by shotgun MS is well below the complete proteome [41, 56]. Indeed, since ions are selected at random for fragmentation and MS/MS analysis, the most highly expressed proteins are identified multiple times at the cost of proteins expressed at low level, which dramatically limits the dynamic range of shotgun MS approaches [41]. A typical shotgun MS experiment offers a dynamic range of detection of 3–4 orders of magnitude, whereas it is estimated that the concentration of proteins can vary up to 10 orders of magnitude in human body fluids [23]. Furthermore, owing to the high-redundancy and extreme complexity of the sample, the full spectrum of peptides present is largely under sampled, which in turn means that repeated analyzes of the same or similar biological samples can show distressingly little overlap of identified proteins [41] since each experiment will sample only a subset of the proteins and not necessarily the same subset in each repeat [56]. To some extent, these problems can be overcome by extensive fractionation and multiple enrichment steps, but this requires an additional non-negligible amount of both experimental and computational work [56].

### 2.2.6 Targeted MS/MS

One way around the limitations of shotgun MS is to adopt a strategy where the mass spectrometer is tuned to analyze specific proteotypic peptides, i.e., peptides that are observable by MS and uniquely identify a target protein. This approach, termed target-driven MS, starts from a list of proteins of interest and carefully selects target peptides for their high propensity to be identified by MS and to uniquely identify a protein or protein isoform of interest. These proteotypic peptides can be identified experimentally (by searching through repositories of observed proteins) or computationally (by predicting them, if the protein has not been previously observed). This type of workflow is called selected/multiple reaction monitoring (S/MRM) and is typically carried out in triple quadrupole type mass spectrometers [41]. The specific proteotypic peptides will be selected in the first quadrupole, then fragmented by collision induced dissociation in the second quadrupole and a second mass filter in the third quadrupole allows for the filtering of the corresponding fragment ions. The identification and quantification of proteotypic peptides is based

on the mass to charge ratios of the precursor and fragment ions pair, which are referred to as “transitions” and are highly specific for a particular peptide [41]. Single reaction monitoring refers to the case where one transition is observed for each peptide, whereas in multiple reaction monitoring, multiple transitions are monitored [41]. In combination with isotope labeling, this technology allows for very accurate, sensitive, and reproducible quantification of the proteotypic peptides that are analyzed. If one can provide the approximate retention time information, then the time of detection of specific transitions can be restricted, therefore allowing for detection of multiple peptides per measurement, a technology referred to as scheduled selected reaction monitoring [41].

Applied to a study of selected proteins in the yeast *Saccharomyces cerevisiae* [56], this technology has been shown to be able to detect and accurately quantify yeast proteins expressed over the full range of cell abundance, from less than 50 copies per cell to over a million copies per cell, without additional fractionation or enrichment steps. This study also demonstrated the capacity of this workflow to comprehensively monitor more than a hundred proteins in a 1 h MS run, which then opens new possibilities for investigating a system under different conditions and replicates. A bottleneck of this workflow, however, is the validation of the SRM transitions that constitute the final mass spectrometric assay in the particular mass spectrometer used for the experiment [56]. Therefore, although targeted MS offers the most sensitive MS detection capabilities to date [23], and unprecedented sample multiplexing capabilities, setting up, optimizing, and validating an assay is relatively time consuming [23, 56]. The accurate mass tag strategy, which is based on the definition (using tandem MS) of peptides whose masses are characteristic of a protein and which can then be detected and quantified by a single MS, can also be used to perform higher throughput targeted MS analyzes [64]. However, this technique suffers from the same drawbacks in terms of time to set up the assay.

### 3 Computational Analysis of Large Scale Phospho-proteomics Data Sets

Having overcome or mitigated all the challenges mentioned above to collect a good quality high-throughput data set, one faces the challenge of interpreting it, which is not a straightforward task and is practically impossible based on inspection and intuition alone. However, mathematical analysis can provide invaluable help in extracting information, that is, not readily apparent. Various approaches to do so are available, and some of them will be described in this section.

We will start by describing applications of methods derived from machine learning and statistics (such as supervised and unsupervised learning, enrichment analysis, etc.). These methods are mainly used for hypothesis generation (i.e., providing leads for areas of further investigation), and usually generate limited

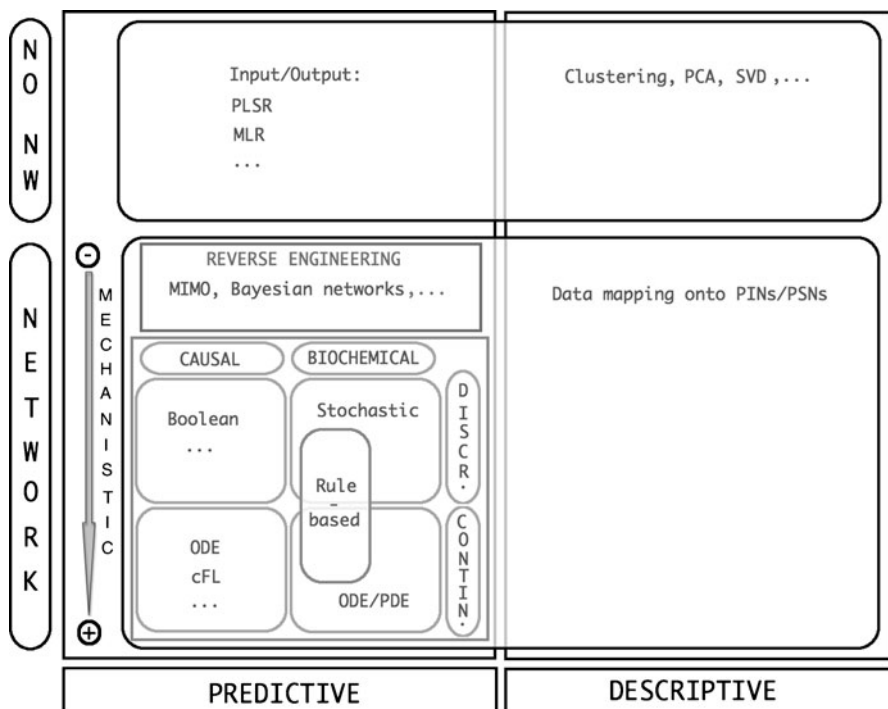
explanatory or mechanistic insights, but they are relatively straightforward to apply to large and noisy data sets. These methods are also generally unbiased (i.e., hypothesis free) and in this sense are a good starting point in an analysis because they provide a good first overview of the data [43], and do not rely on extensive a priori expert knowledge which might not even be available for the system under investigation.

Another set of approaches that is frequently applied in the same context as the above, is the mapping of data (e.g., differential expression/modification, phenotypic data, etc.) to known or derived “pathway maps.” All of these approaches are very familiar to the field of functional genomics and in that sense these methods are quite mature and well known. These types of analyzes have recently been applied quite extensively to investigate large scale phospho-proteomics MS experiments in various settings, which is what we will discuss in the first part of this section. We will refer to these methods as “descriptive” approaches.

Although the methods mentioned above have the potential to generate useful hypotheses, they do not address a fundamental functional characteristic of signaling systems, which is the ability to process information (input) and produce a response (output). To study this process, we need to generate more detailed and hopefully more realistic models of what happens in the cell when a signal is processed. Such models include, but are not limited to, partial least square regression, ordinary and partial differential equations (ODE/PDE), Bayesian networks, rule-based, and logic-based models. We will refer to these formalisms as “predictive approaches”. These models are predictive in the sense that given a set of conditions that was not present in the data used to build the model, they should be able to predict the behavior of the system. These methods usually generate explanatory and mechanistic hypotheses (although the actual mechanisms are described with broadly variable levels of details and therefore so are the insights generated, so care should be taken when interpreting them). There are many ways to look at and classify different types of modeling approaches, and all of them are somewhat artificial. The distinction that we make between “descriptive” and “predictive” models is only made for organizational purposes and is not intended as an absolute or universal classification (Fig. 2.2).

### 3.1 “Descriptive” Approaches

In this section, we will describe some “data-driven” approaches to signaling networks that have been applied to MS and affinity based large scale phospho-proteomics data sets, and briefly mention some of the insights that have been extracted from these analyzes. Whereas affinity-based data sets are now extensively used to generate complex quantitative models, MS proteomics data sets are mainly still at the stage where descriptive investigations are a necessary first step to make sense of the wealth of information that is generated.



**Fig. 2.2** Model formalisms that have been applied to signaling networks. Although there are many ways to look at and classify these methods, we chose to distinguish between “descriptive” and “predictive” approaches, on the basis that the latter allow prediction of a system’s response given a set of conditions and data on predictor variables. Within these categories, we further distinguish between methods that rely on or produce a network (i.e., a graph composed of nodes and edges), and methods that do not. In the group of predictive network-based approaches we can further distinguish between approaches that do not necessarily rely on previous knowledge about the system’s connectivity (reverse engineering), and methods that rely on some sort of previous knowledge. The latter group can be further divided into 4 categories depending on their discrete/continuous nature and on their causal/biochemical character. Causal approaches only seek to determine relationships between species (such as protein A activates protein B) whereas biochemical approaches include some degree of mechanistic description of the reactions at play. PLSR = partial least square regression, MLR = multiple linear regression, PCA = principal component analysis, SVD = singular value decomposition, MIMO = multiple inputs multiple outputs models, ODE = ordinary differential equations, PDE = partial differential equations, cFL = constrained fuzzy logic, PINs = protein interaction networks, and PSNs = protein signaling networks

### 3.1.1 Global Investigations of the Phospho-proteome

In view of the highly complex task of making sense of high-throughput phospho-proteomics data in a signaling context, several tools have been developed specifically for this type of data, such as PTMscout [50], NetworKIN [40] or the PHOSIDA [21] database, amongst others. PTMscout is a web-based interface for viewing,

manipulating, and analyzing high-throughput PTM data. Analysis capabilities focus on hypothesis generation through subset selection and enrichment analysis based on annotations (such as GO, Pfam, local sequence features, etc.) or user-defined criteria on dynamic profiles [50]. This tool also provides help in the assignment of peptides to proteins by providing orthogonal information such as annotations and mRNA expression when available.

NetworkKIN [40] is an algorithm for prediction of kinases from experimentally determined phosphorylation sites, that integrates sequence specificity with contextual information extracted from resources such as interaction and pathway databases, literature mining, mRNA expression studies, etc. The improved accuracy of this prediction algorithm compared to non-contextual versions indicates that effects such as subcellular compartmentalization, anchoring proteins, temporal, and cell specific expression, etc., play a crucial role in determining kinase-substrate specificity. This in turn points again to the fact that signaling is a highly context specific concept, and that a network level understanding of kinase activity is likely to be necessary even when it comes to understanding single molecular events.

PHOSIDA [21] is a phosphorylation site database for large scale and high confidence quantitative phospho-proteomics experiments that allows the retrieval and analysis of such data, and includes information on evolutionary conservation as well as a phosphorylation site predictor. Other databases, such as the manually curated phosphorylation site database PhosphoSite [26], offer additional information such as association with diseases and sequence logos. The databases mentioned above include some type of analysis tools, but there are also other data repositories that can be valuable resources for proteomics investigations, such as the PRIDE [72] and Phospho.ELM [15] databases.

An interesting perspective on the global function and investigation of phospho-signaling was recently provided by Bodenmiller [6]. In this study, 97 kinases and 27 phosphatases in yeast were systematically knocked out or inhibited, followed by phosphopeptide enrichment and label free LC-MS/MS identification of more than 1,000 phosphopeptides showing a significant change in abundance compared to a wild type situation. Analysis of the direct versus indirect effects of these deletions led to the observation that not a single kinase showed exclusively direct effects. Furthermore, analysis of growth speed and morphological features of each deletion strain revealed that the phenotype strength was not necessarily reflected in the magnitude of the effect on the phospho-proteome. Together, these observations reinforce the view that signaling has to be very flexible and redundant to allow the cell to respond to a changing environment, and point to the fact that modulating any branch of a network might not be possible without system-wide adaptations.

### 3.1.2 Analysis of Pathway Utilization Downstream of Receptors

The study performed by Olsen [54] set the stage for MS analysis of signaling by pointing both at the complexity of the problem and the sparseness of our knowledge of the involvement of phosphorylation in signaling. Using a strategy



combining phosphopeptide enrichment, high-accuracy identification by LC-MS/MS and SILAC, they were able to quantify dynamic changes in phosphopeptides levels at 6,600 sites on 2,244 proteins upon stimulation of HeLa cells with EGF for different times (from 0 to 20 min). In addition to this, this study also includes some spatial information since nuclear and cytosolic fractions of each condition were obtained and analyzed. Using a cutoff of a minimal 2-fold change upon stimulation, the authors determined the sites that were dynamically regulated and performed fuzzy *c*-means clustering (where each point belongs to clusters with a certain degree, depending on its distance to the centroids of the clusters) to identify groups of sites with similar dynamic profiles. The main conclusions of this study included the observation that most *in vivo* occurring phosphorylation sites had probably not been detected before, and that groups of phosphosites from the clustering analysis do contain functionally related members.

Another important result of this analysis was obtained by looking at phosphosites which map to the same protein. Indeed, the authors observed that 77% of proteins that had a regulated phosphopeptide also had at least one other site whose regulation profile was different (either unchanging or belonging to a different cluster of the above analysis). This underscores the fact that when looking at the degree of phosphorylation of proteins we should always measure site specific events if we want to obtain accurate and functionally relevant information. This also points to the complexity of interpreting phosphorylation data since it seems that phosphorylation can serve different functions at different sites in the same protein. Finally, Olsen et al. [54] noted that only a subset of the proteins found to be dynamically regulated by EGF signaling were known to be involved in growth-factor signaling, which points to potential gaps in our knowledge of even well-studied pathways.

A similar system was investigated by Huang et al. [27] with different goals and methods, with the objective of determining differences in the signaling downstream of a truncated extracellular mutant of the EGF receptor (EGFRvIII, frequently found in glioblastoma multiforme), compared to the wild type EGFR, and depending on the level of expression of the mutant receptor. The workflow of this analysis was as follows: transduced U87MG glioblastoma cell lines expressing differential levels of EGFRvIII were isolated by FACS, peptides from these cell lines were then isolated, stable isotope labeled and mixed. Next, tyrosine phosphorylated peptides were immunoprecipitated and further enriched by IMAC, and finally analyzed by LC-MS/MS. Quantitative phosphorylation profiles were generated for 99 sites on 69 proteins, which were mapped to canonical EGFR signaling cascades. This indicated that signaling downstream of EGFRvIII and wild type EGFR favour different routes, and this is also dependent on the level of expression of EGFRvIII, e.g., cells that highly overexpress EGFRvIII preferentially use the PI3K pathway over the MAPK and STAT3 pathways. Using a self-organizing map, the authors also identified phosphotyrosine sites with similar profiles, which led to the identification of a cluster of sites that significantly increased as a function of EGFRvIII expression. Examination of the members of this cluster led to the hypothesis that the EGFRvIII receptor was constitutively activating the cMet pathway. Finally, quantification of the phosphorylation sites on the receptor itself pointed to differences in regulation



between wild type and truncated receptors. Altogether these observations indicate that although phosphorylation of the EGFRvIII might not be qualitatively different from the wild type situation, quantitative differences at each individual site might have functional implications reflected in different utilization of downstream pathways and therefore different biological responses [27]. This in turn means that a quantitatively accurate model of this system is likely to prove very useful.

The paper by Krueger et al. [37] aimed at determining the tyrosine phosphoproteome of the insulin signaling pathway by stimulating SILAC labeled differentiated brown adipocytes with insulin for various times, then immunoprecipitating phosphotyrosine containing peptides and analyzing them by LC-MS/MS. Thirty three proteins were identified to be significantly regulated, which was confirmed by western blot. By looking at the dynamic profiles and fold activation of the proteins in this candidate list, they were able to generate hypotheses for new insulin induced candidate effectors and to link them with branches of the insulin pathway.

Matsuoka et al. [45] also used MS to investigate phosphorylation events downstream of a cellular signal, this time concentrating on the landscape of the DNA damage response (DDR) mediated by the ATM and ATR kinases. Briefly, they mixed and immunoprecipitated peptides from two SILAC labeled populations of HEK 293T cells, one having been exposed to ionizing radiations, using antibodies to phospho-SQ or phospho-TQ (ATM and ATR recognise Ser-Gln and Thr-Gln motifs). The samples were then subjected to LC-MS/MS and 905 phosphorylation sites on 700 proteins were shown to display a more than four fold increase following DNA damage by ionizing radiation. This list of proteins was then examined manually, and mined for enriched GO annotations and functional modules using the softwares from Ingenuity. This showed an enrichment for proteins involved in nucleic acid metabolism, and revealed many clusters of proteins previously known to be interacting, but not necessarily known to be involved in the DDR. A subset of the proteins in this list, that were not previously known to be involved in the DDR, was also examined for functional involvement in this response using siRNAs. Although the approach applied here cannot formally distinguish between direct targets of ATM and ATR kinases and targets of kinases with similar specificity, all identified phospho-sites are likely to be regulated by the DDR, and their belonging to a large number of interconnected functional modules suggests an impact of the DDR on cellular physiology that is far broader than expected [45].

More recently, phospho-proteomics was again used to generate qualitative hypotheses using pathway enrichment, this time with the objective to investigate signaling events downstream of the mutant protein NPM-ALK, which is common in positive anaplastic large cell lymphomas [76]. GP293 cells were transfected with either NPM-ALK or a NPM-ALK mutant with decreased tyrosine kinase activity (used as a negative control), the phosphopeptides were then purified and subjected to LC-MS/MS. This led to the identification of 506 phosphoproteins present only in NPM-ALK expressing cells, from which a pathway enrichment analysis was performed (using a Fisher exact *t*-test). The samples were also hybridized to antibody arrays and differential phosphorylation was used as a basis for pathway

enrichment. Both methods resulted in a substantially overlapping list of enriched pathways, from which the authors chose to focus on the TNF/Fas/TRAIL pathway, performing various validations of the involvement of this pathway (comparison with a list of previously generated potential binding partners of NPM-ALK, western blot quantification of three proteins in this pathway, and siRNA knock down of two of those, combined with a viability assay of the knock down of TRAP1 upon drug treatments). This study again underlines the ability of MS data to generate qualitative hypothesis regarding signal transduction.

### 3.1.3 Analysis of Reciprocal Signaling in Cell–Cell Communication

The approach adopted by Jorgensen et al. [33] is slightly different and addresses a fundamental biological fact, that is, signaling usually happens within the context of tissues and often involves multiple populations of cells. This is particularly important when the signaling is initiated by cell–cell contact, as in the case of the ephrin–EphR interaction. In such cases, the signaling typically involves the reciprocal exchange of distinct information between the interacting cells, leading to mutually coordinated alterations in their respective behaviors. Therefore, stimulating such systems with soluble versions of the ligands is an artificial setting that might provide only limited understanding (e.g., signaling between EphR and ephrin expressing cells might be influenced by interactions with adhesion molecules).

Therefore, in this study, EphB2 and ephrin-B1 expressing populations of HEK293 cells were SILAC labeled and co-cultured for 10 min, then lysed and mixed with non-stimulated EphB2 expressing cells as a reference, before phosphotyrosine peptide isolation and LC–MS analysis. This led to the identification of 442 sites on 304 proteins that significantly decreased or increased in abundance upon stimulation, in one or both cell types, revealing common and cell specific modes of regulation. The authors then turned to a siRNA screen in which monitoring of the cell sorting response when mixing the two cell populations (when mixed, EphB2, and ephrin-B1 expressing cell populations form distinct colonies with well defined boundaries) allowed them to propose a list of proteins involved in this phenotypic response. Using the NetworKIN [40] and NetPhorest [46] tools, a network was constructed based on the prediction of kinases, phosphatases, and phospho-binding modules for each phosphotyrosine that was found to be modulated upon cell–cell contact. These predictions were then pruned based on criteria from the MS and siRNA analyses, and other information such as protein interactions. The obtained network was then represented in a cell-population specific way using the modulation of phosphotyrosine sites determined by the MS analysis. Finally, the MS experiment was repeated using a variant of ephrin-B1 that lacked the cytoplasmic tail, thereby impairing its ability to relay the signal inside the ephrin-B1 expressing cells, but not its ability to interact with EphB2. A significantly different response was observed in the EphB2 expressing cells in this case compared to when the full ephrin-B1 was used, thereby confirming that there is a bidirectional signaling process at play in the

system. This study not only demonstrates the power of MS to investigate complex signaling systems, but also points to the limitations of the in vitro systems in which we commonly conduct our investigations.

## 3.2 “Predictive” Approaches

In this section, we will describe more detailed and predictive approaches to modeling of signaling networks that have been applied to proteomics data sets, mostly acquired using affinity based technologies. By “predictive” we mean that these models are often capable of computing the expected state or evolution of the system when under particular conditions (e.g., when applying an inhibitor against one of the species in the model). We will start with simple (linear) regression based models that can predict some variables based on linear combinations of other measurements. We will then briefly touch on other correlation based methods. This will be followed by the presentation of ordinary differential equations (ODEs) as a natural way to describe processes where species of interest are changing as a function of time in a quantifiable manner [5]. Then, in light of the extraordinary combinatorial complexity that often arises in signaling systems, we will discuss alternative methods to model detailed signaling networks, such as logic-based and rule-based approaches. Finally, we will discuss the role of previous expert knowledge in the inference process and briefly present Bayesian networks as a strong statistical approach to deal with this.

### 3.2.1 Input/Output Regression Based Approaches

Two linear regression based approaches will be described here, partial least squares regression (PLSR) and multiple linear regression (MLR). In PLSR, the data are separated into a set of inputs and a set of outputs, which are then reduced to their principal components and a linear solution is identified that relates the inputs to the outputs. PLSR can be used to determine which inputs display the biggest correlation with outputs for example, and it can also be used to predict the outputs from inputs measured in new experiments. MLR is similar to PLSR but the linear solution is computed directly between the measured variables, without dimensional reduction, which makes its results easier to interpret [2]. However, both models suffer from the same limitation: being linear models, they cannot capture coupled effects and nonlinear phenomena such as saturation, switch like effects, etc. [51].

MLR can and has been used to reconstruct network topology from experimental data, for example, in the study by Alexopoulos et al. [3]. In this paper, primary hepatocytes and HepG2 liver cancer cells were exposed to multiple conditions made of combinations of one of 7 growth factors or cytokines, in the presence or absence of 7 small molecule kinase inhibitors. The level or state of modification of 17 intracellular proteins and 50 secreted peptides were measured using a

sandwich immune assay with the xMAP platform. MLR was then performed to relate signals to cytokine secretion, and to relate cues and inhibitors to signals. The regression weights were then used to draw connections between ligands and readouts which allowed the comparison of immediate-early signaling downstream of 7 transmembrane receptors in normal and transformed hepatocytes [3]. Edges selected based on greatest differential regression weights between hepatocytes and HepG2 cells were selected for further experimental investigation. From this analysis, the authors were able to conclude that the magnitude of responses to stimulations (whether reflected in the intracellular signals or in cytokines secretion) were vastly different between the two cell types and that even when both cell types are responding to the same ligand, the extent to which specific downstream pathways are activated is very different.

In the work by Gaudet et al. [20], PLSR was used to extract information from a vast compendium of data acquired from multiple assays such as kinase activity, quantitative immunoblotting, and antibody microarrays. Briefly, HT-29 cells were treated with  $\text{TNF}\alpha$ , in combination with EGF or insulin, and 19 protein signals were measured over 24 h, along with 4 different measurements of apoptotic response measured by flow cytometry. PLSR is then used to relate signaling data to apoptotic responses. The authors showed that the model derived from the full compendium and a set of metrics derived from the time course data performed extremely well when assessed by leave one out cross validation and independent validation on a new data set. The authors also showed that models built on single protein measurements were poorly predictive, and more surprisingly that models built on measurements of multiple signals from single types of assays were also inferior. Furthermore, models built from the raw measurements only performed poorly on the validation data set, whereas models built only from the derived metrics capturing the time dependent profiles of the signals performed as well as the full model. This points to the fact that time-dependent information is crucial to the predictive power of the model. Finally, they showed that models based on data obtained with cells exposed to multiple combinations of cytokines are less sensitive to experimental noise. In the related study by Janes et al. [29], the contribution of single proteins to the apoptotic response was investigated, and the proteins JNK1, MK2, and ERK were found to provide the most information for prediction of the apoptosis status, based on the average information contained in their derived metrics. The authors also noted that prediction efficiency was maximal with 4–5 signals, and that a model derived from signals measured only in the first 4 h after stimulation (before the onset of apoptosis) were already sufficient to predict the apoptotic signature.

### 3.2.2 Network Inference

Many methods to build models of signaling networks rely to some extent on previous knowledge about the system under investigation (e.g., a fully detailed mechanistic description of the process at hand in mechanistic models, or a simple description of the logic interactions involved in logic-based models). Building such

models involves a literature (or database) search which is not only usually heavy in terms of workload, but it is also error prone because many molecular events are context specific and the context of an interaction is not necessarily reported. This also biases investigations towards well studied systems. Some formalisms however allow the reconstruction of signaling networks entirely from data, without relying on any type of mechanistic knowledge.

The regression based methods presented above require only very limited prior knowledge, i.e., determining which variables are dependent and which ones are assumed to be explanatory. These methods do not rely on any graph (network) structure, but only predict some variables based on their statistical dependency with others. Interestingly, in the context of the DREAM initiative ([www.the-dream-project.org](http://www.the-dream-project.org)) when confronted with the challenge of predicting unseen measurements of proteins and/or cytokines for combinations of stimuli and inhibitors of a signaling pathway (based on measurements of the same players under different combinations of the same stimuli/inhibitors), methods that performed the best used a statistical approach that did not rely explicitly on an underlying signaling network [57]. Duvenaud et al. [16] also reported that functional causal models that predict the effects of actions on the system (as conditional density models) without relying on any graph tend to perform well or better than methods for learning conditional density models based on graphs. There are many other correlation based methods that can be applied to signaling networks, and many of those have been developed for gene regulatory networks (see the following for reviews [4, 44]). One should be aware, however, when interpreting such analyzes that a correlation does not necessarily mean a causal link, and that correlations can encompass both direct and indirect interactions.

In Ciaccio et al. [11], for example, the algorithm ARACNe [42], which was originally developed for microarray expression profiles, is applied to the analysis of a data set on 91 phosphosites on 67 proteins at 6 time points after stimulation with 5 EGF concentrations, obtained using microwestern arrays. The algorithm uses information theoretic approaches to prune indirect interactions inferred by co-expression methods. In Santos et al. [63], an approach called modular response analysis [36] is used to determine the MAPK network architecture in the context of NGF and EGF stimulation. This method is a sensitivity analysis based process relying on measuring network responses to successive small perturbations (here implemented by RNAi), at steady state conditions. Network connections are inferred by computing local response coefficients, which estimate the sensitivity of one module of the network to perturbation of another module, in isolation of the total network [63]. Although in this analysis the system studied is much smaller than those interrogated using high-throughput proteomics, similar approaches could be used to study larger systems.

In the work by Nelander et al. [51], a methodology is proposed to derive network models from time courses of evolution of molecular species upon perturbations. This works builds upon the type of models called multiple inputs multiple outputs (MIMO) models where the time dependent evolution of activities of the system's components (outputs) are described by differential equations as nonlinear functions

(transfer functions) of themselves and a vector of perturbations (inputs). Within the nonlinear transfer function the dependencies between elements of the system are described as linear combinations of the components. The coefficients of these linear dependencies can be interpreted as strength of interactions between the nodes, assuming that they reflect underlying causal relationships between the components, thereby making it possible to derive a node–edge representation of the inferred system (where an edge is present when this strength of interaction is above a certain level) [51]. This representation (as any purely data-driven) has the disadvantage that the nodes in the model are the perturbed and observed molecular species only, which might not be identifiable as single molecular species, and it ignores any unperturbed and unobserved species that might be involved in the connectivity structure of these nodes [51].

### 3.2.3 Bayesian Network Inference

Bayesian networks have the natural ability to accommodate previous knowledge to a chosen extent. Depending on the level of information that one wants to put in the prior of the models (see below), one can make the inference process entirely independent of any prior knowledge (flat prior) or bias the inference towards models that are casted “more likely” based on a priori expert knowledge. A Bayesian network consists of a directed acyclic graph with vertices representing the molecular species to be modeled as random variables, edges describing conditional independencies between those variables, and parameters describing the conditional distributions implied by the graph (e.g., when the states are discrete, this typically takes the form of a probability for a target node to take each of its possible states given all possible combinations of states of its parents nodes). The graph structure implies that each variable is conditionally independent of all non-children nodes given its immediate parent nodes [49]. Bayesian network inference aims at making inferences regarding the structure of the graph using Bayes’ theorem, which states that the posterior probability of a graph (probability of a graph given the data) is given up to proportionality (i.e., ignoring a normalizing constant when comparing structures obtained from the same set of data and distributional assumptions) by the product of the marginal likelihood (probability of the data given the graph) and the prior distribution over directed acyclic graphs (i.e., how likely is each individual graph structure) [49]. Using certain distributional assumptions, the posterior probability of graphs can be computed up to proportionality, which is enough to compare graphs in a search procedure, in order to find a graph structure that is optimal under the statistical model at hand.

Given their solid basis in statistics, Bayesian networks are naturally able to handle stochastic aspects of biological processes and noisy measurements [31]. However, this comes at a high cost in terms of data requirements. Such an approach is, however, ideal when the data at hand is cell specific and therefore each measurement includes data about a whole population of cells at the single cell level, as is the case in the study by Sachs et al. [59] where intracellular multicolor

flow cytometry is used. In this paper, Bayesian network inference is used to investigate signaling networks of human primary naive CD4+ T cells, downstream of CD3, CD28, and LFA-1 activation, based on measurements of 11 phosphorylated proteins and phospholipids [59]. Similarly in [11] Bayesian networks are used to model the dependencies between 67 proteins (measured by microwestern arrays at 6 time points) after stimulation with EGF at 5 different concentrations. In this case, in order to have enough samples for each measurement, each time point and concentration of the stimulus is used as an independent sample, yielding 20 samples per measurement. One of the strength of Bayesian networks in this context is that, using carefully chosen prior distributions on graphs, it is possible to include information on network features such as particular edges, types of edges, degree distribution, and sparsity. In practice this means that not every possible graph is considered equally plausible, and that we can bias the search towards graphs that we consider a priori more likely [49]. This in turn has the advantage of constraining the space of possible graphs to search, which makes the inference process more efficient, while maintaining the Bayesian networks' natural ability to deal with noise and stochasticity. The expert knowledge involved in specifying those priors can be as detailed as specifying a particular edge to be very likely or as vague as specifying that ligands should generally interact with receptors and not effectors [49].

When interpreting Bayesian networks it is important to be aware that many Bayesian networks can represent the same statements of conditional independence, i.e., the inference process can be unable to distinguish among a series of graph with the same undirected graph but in which some edges might have different directions [44]. However, perturbing the states of measured molecules with molecular interventions can help resolving this problem by providing information on the causal relationships between nodes [59]. Furthermore, a limitation of Bayesian networks is that they are constrained to be acyclic, which means that feedback loops for example cannot be uncovered. However this limitation can be overcome by using dynamic Bayesian networks [59].

### 3.2.4 Reaction-based Models

All of the models described above infer a topology as statistical dependencies between variables, not mechanistic links. If some mechanistic knowledge about the topology of the system is available, then other methods can be applied that incorporate this information. An extreme case compared to network inference is the application of ODE/PDE models where detailed knowledge about the the biochemistry (reactions) of the system is written down as a set of differential equations.

Biochemical (also called physicochemical [1]) models describe the temporal evolution of individual biomolecular species as functions of their rates of production and consumption in terms of mass action kinetics, which is an empirical law expressing the rates of reactions as proportional to the concentrations of their reactants [1]. In the simplest case one uses ODEs, and spatial heterogeneity (i.e., changes



in the location of species) is represented by compartmentalization, where each compartment is assumed to be perfectly well mixed (i.e., instantaneous transport inside a compartment, leading to homogeneous concentrations of all species across the whole compartment). Partial differential equations (PDEs) arise when the spatial dimension is explicitly modeled, i.e., spatial gradients are now included in the representation. Building an ODE/PDE model involves three main steps, often applied in an iterative way: model development (write down biological knowledge in terms of rate of change equations), parameter estimation (determining the values of unknown parameters), and model validation (comparing model predictions to independent experimental data) [5].

When designing the models, two critical decisions need to be made: what is the scope of the model and at which level of detail will the system be described [9]. Defining the scope involves determining how much of the system needs to be modeled in order to achieve the goal of the modeling process, and deciding on the level of detail involves choosing a level of representation of the molecular species and complexes (i.e., do we want to represent all modifications and interactions explicitly). The latter point is especially challenging because biological species are often capable of assembling into multi-component complexes, undergoing multiple PTMs, and segregating into various sub-cellular compartments and locally concentrated areas, and we often do not know how to interpret these events in terms of signals. This latter problem is referred to as “combinatorial complexity” and is what quickly makes ODE and PDE models untractable [7, 12, 17, 18]. Another common problem with ODE/PDE models is parameter estimation, which involves determining the range of parameter values over which the model closely reproduces the experimental data [1]. Problems arise in this process when the model reproduces the experimental behavior equally well over a large range of parameter values, therefore making those parameters unidentifiable.

Some common simplifying assumptions are made to overcome the problem of combinatorial complexity, such as ignoring intermediate states of assembly when they are fast, or lumping together biochemical forms that are thought to be equivalent. However, these remain assumptions, and just as any other assumption made in building the model (e.g., well mixed compartments, etc.) it is very important to be aware that the equations obtained are only valid given all of the assumptions made, and so each assumption and the implications thereof should be discussed, in light of explicitly stated design goals [1, 5]. It is also important to note that ODE and PDE models are deterministic continuum approximations of what happens in the system [9]. When limited number of molecules are involved in a process (e.g., small compartments or slow reactions), then stochastic effects may become important and a deterministic approximation might not be able to accurately represent the evolution of the system [1].

Despite these complications, ODE and PDE models can be used to generate valuable insights into biological questions. In Birtwistle and Kholodenko [5], for example, the authors describe how simple and more complex ODE and PDE models can be used to gain insights into the role of endocytosis in signaling. In the paper by Chen et al. [9], for example, a detailed model of ERK and Akt



regulation by two ErbB ligands and four ErbB receptors during the immediate-early phase of ligand stimulated cell signaling is built, parameterized, and analyzed. This model includes 28 proteins, but accounting for protein–protein interactions, PTMs, and compartmentalization generated an additional 471 species, requiring 499 differential equations, 201 unique reaction rates, and 28 non-zero initial conditions. This leads to a complex parameter optimization problem, despite some parameters being measured or extracted from the literature. Other parameters were estimated from the data by minimizing the difference between experimental and simulated data [9].

This model was found to be unidentifiable, with some parameters being quite constrained across similarly performing models (in terms of fit to data), and some parameters spanning the entire range of values allowed in the search. However, the authors were still able to perform a sensitivity analysis of the partially calibrated models (i.e., an investigation of which parameters have the largest influence over a chosen observable, when varied), as well as a dose responsiveness analysis, and to extract useful predictions from those analyses. For example, they showed that the calibrated  $v_{\max}$  for the PP2A compartment targeting pRAF and pMEK was markedly different from the compartment targeting pAkt, which led them to hypothesize that dephosphorylation of Raf, MEK, and Akt occurs at different rates, and that this presumably involves different PP2A-containing complexes. The sensitivity analysis also yielded valuable insights, such as which parameters have the biggest influence on EGF- or HRG-stimulated pERK across multiple partially calibrated models, and the observation that parameter sensitivity critically depends on the observable that is chosen [9]. This shows that, provided that care is taken in interpreting the results of an analysis, and that parameter uncertainty is considered in this process, even partially calibrated models can provide valuable insights.

### 3.2.5 Rule-based Models

A formalism that naturally describes the mechanisms of signaling systems despite their associated combinatorial complexity is the principle of rule-based model. A rule-based description of a system allows a rich variety of knowledge about this system to be expressed in a single formalism (see [25] for a review) [31]. Briefly, the system is described as a set of agents which have labeled sites that can each have an internal state, typically used to denote PTM status. The agents are acted on by rules, which provide descriptions on how they interact, with common interactions consisting of binding/unbinding of agents, modification of the state of a site, and deletion/creation of an agent. The left hand side of a rule specifies a condition that applies on a pattern of agents and their site values, whereas the right hand side specifies actions on agents mentioned on the left. Only the information that is triggering the accomplishment of the rule needs to be specified on the condition side of the statement [14].

Simulation of a rule-based model can be performed by the repeated process of matching the facts (patterns of states of agents) against the condition part of

the rules and carrying out the action part of the rules where the condition part is satisfied [31]. A control strategy is used to determine the order in which the rules are applied, which typically takes the form of a rule-based version of Gillespie's algorithm [14, 31, 65]. Popular languages to write and simulate such models are Kappa [14], and BioNetGen [17] which is extended in the software NFsim [65]. Differential equations can be also derived from the rules; if all possible species are described the number of states increases exponentially due to the combinatorial explosion, but methods exist to simplify them at least to some degree [7, 12, 18].

### 3.2.6 Logic-based Models

Whether as reactions or rules, building a biochemical model requires a lot of mechanistic information about the system, and the resulting models are difficult to simulate. This limits their applicability to relatively small and well studied systems. However, data generated by high-throughput methods typically provide wide scope information which leads to the need for formalisms capable of handling big networks for which only limited mechanistic knowledge is available. A suitable formalism to model large networks for which some mechanistic knowledge is available are logic-based models, that include dependencies between components, while ignoring the molecular details [31, 47, 74]. In logic-based models dependencies between nodes are specified in terms of gates, which are associated with truth tables that describe output states for all possible combinations of input states [47]. If two proteins A and B have a positive effect on the activation of a third one C, the corresponding gate can be either an OR (either A OR B activates C) or AND (only A AND B together activate C).

The simplest type of logic model is a Boolean model, in which each state is either on or off (1 or 0). Following the pioneering work by Kauffman [35], Boolean logic models have been used extensively to model genetic regulatory and signaling networks [31, 47, 74]. This formalism allows one to compute the state of activity of each node of a graph given different inputs or initial states. Cause-effect relationships in biological pathways can often be found in the literature, and in databases such as reactome (<http://www.reactome.org>) or panther (<http://www.pantherdb.org>). However, these resources rarely include specific gates, nor cell-type specific information. This problem can be overcome by using signaling data to train a Boolean model from a generic prior knowledge network derived from the literature or databases [61]. By pruning the network, one obtains models with a much higher predictive power, that are specific to the data (and thus cell-type) they have been trained to. Thus, by leveraging prior knowledge and dedicated signaling data, one can model relatively large networks with relatively sparse data, and because one includes intermediates (not just perturbed or observed variables), the mechanistic insight is higher than in purely data driven models. Thanks to their simplicity, these models can easily accommodate  $\sim 100$  nodes and be trained to phospho-proteomics sets of  $\sim 1000$  data points [61].

A main limitation of the boolean logic approach is that all species are considered either on or off, and the model is therefore not able to account for intermediate levels of activation. Fortunately, several logic-based extensions provide a means to do so, such as multi-state discrete models and fuzzy logic [47]. In multi-state discrete models, additional levels between 0 and 1 are specified, whereas in fuzzy logic a set of user-defined functions are used to transform discrete logic conditions into relationships between continuous inputs and outputs. An extension of the approach from [61] was recently proposed that allows the training to data of a fuzzy logic model obtained from previous knowledge [48]. The approach is termed “constrained fuzzy-logic” because the set of relationships between model species is limited, thereby making it possible to train both the topology of the network and the particular quantitative relationship involve at each gate, and allows to model features not captured by Boolean logic [48]. However, this ability comes together with an increase in complexity that renders the approach more difficult to apply to large networks (above a few dozen species) [48].

Both of the approaches described above compute a steady state of the logic model. However, logic-based models can integrate the notion of time, with various degrees of detail. To compute a trajectory of the system, the status of nodes are updated (as functions of the state of their input nodes) at each (time) step according to two main updating schemes: synchronous, where all nodes are updated simultaneously with a new state depending on the state of each node’s inputs at the previous time step, and asynchronous, where nodes are updated in random order with a new state depending on the state of some input nodes at the previous and some at the current time step [47]. Mixed asynchronous schemes allow some nodes to be updated before others, making it possible to model separate time scales. Logic models can also be converted into ODEs, making both species and time continuous, albeit at the cost of increased complexity [47].

Finally, logic-models can be extended to incorporate probabilistic interactions, thereby incorporating uncertainty in biological knowledge and/or stochasticity of the system [47]. Logic-based models can also be implemented in a Bayesian framework (see [19] and [39] for more information).

## 4 Summary

Modeling is an invaluable tool to make sense of large and/or complex systems from a functional perspective. Signal processing involves regulations on the proteome at three highly regulated and coordinated levels: regulated post translational modifications (PTMs), protein–protein interactions, and changes in expression levels. The PTM level is what we focus on here because it is the most immediate one and often triggers changes at the other levels, and in particular we concentrate on phosphorylation as a major regulator of protein function and activity. Many proteins are modified at many sites in a highly dynamic and context dependent manner, and combinations of modifications can have various functional consequences that we

are only beginning to unravel. Therefore, interpreting PTM data from a signaling perspective is still a significant challenge, and investigations in this area are likely to benefit from modeling approaches.

A modeling process requires a data set and an appropriate modeling framework. Tables 2.1 and 2.2 summarize the main features of particular applications of modeling pipelines that have been mentioned throughout this chapter. Antibody based approaches allow the quantitative measurement of protein or protein modification levels using technologies such as protein arrays, reverse phase arrays, xMAP, intracellular multicolor flow cytometry, and microwestern arrays. All of these platforms have different limitations in terms of samples and targets multiplexing, signal to noise ratios, and dynamic range. However, they are all based on the recognition of a target by a specific antibody and therefore all suffer from the same limitation: the availability and quality of antibodies. The ability of these methods to be applied to many samples in parallel is a significant asset because it allows for multiple perturbation experiments that inform the network inference process. Compared to MS approaches, antibody based technologies offer limited protein coverage but are more easily scalable to large number of samples [2, 62], and in general require a smaller amount of sample [11].

In contrast, the classical mass spectrometry workflow (shotgun MS) is a non-biased approach (i.e., it is not aimed at particular proteins) that allows the detection of many more proteins in a single experiment. The unit that is identified in LC-MS/MS workflows is a peptide, and peptide mapping to proteins is a non-trivial problem, especially in higher eukaryote organisms where a high level of sequence redundancy can be expected, thus the importance of rigorous statistical approaches for assessing protein identification. Shotgun MS is inherently biased towards peptides that are highly abundant and easy to detect, and selection of peptides for MS/MS is a random process that undermines the reproducibility of shotgun LC-MS/MS approaches. Shotgun MS is somewhat limited with regards to the number of samples that can be processed. Targeted MS is likely overcome some of these limitations, since this technology has the ability to be highly quantitative and reproducible, with an unprecedented dynamic range and the ability to investigate many conditions. However, this method requires a long time for workflow optimization, which means that a significant workload investment has to be done before being able to collect data. With the advances in instrumentation and the emergence of targeted proteomics workflows, MS now has the potential to represent a viable and a more powerful, fully quantitative alternative to antibody based methods. Therefore we expect MS to play a crucial role in the field of modeling of signal transduction networks in the future, provided that modeling frameworks are adapted to the particular features of such data sets.

Whatever the method used, it is important to systematically document and report the pipeline that is used from the data collection to modeling (e.g., normalization in the case of antibody based methods, peptide identification, protein inference, and quantification in the case of MS). Ideally, both the raw and processed data should be available alongside detailed methods for any reinvestigation or even reinterpretation of results. This is particularly challenging in the case of MS since

**Table 2.1** Overview of applications mentioned in this chapter using mass spectrometry

Paper	Goal	Technology	Model	Conditions	Readouts
Bodenmiller et al. [6]	Determine influence on yeast ppome of systematic deletion of kinases and ppase	TiO <sub>2</sub> ppide enrichment, LC-MS/MS	DE, correlation impact on ppome/orthogonal phenotype	124 deletion/analogue-sensitive strains + wild type = 125	8814 peptides, 1026 proteins
Olsen et al. [54]	Phosphorylation events downstream of EGF stimulation	SILAC, TiO <sub>2</sub> ppide enrichment, LC-MS/MS	D.E., clustering	5 tp (stimulation time), 2 subcellular fractions	6,600 sites, 2,244 proteins
Huang et al. [27]	Differences in signaling downstream of EGFR mutant	Stable isotope labeling, TyrP-IP, IMAC, LC-MS/MS	Mapping on pathway, clustering	4 cell lines	99 sites, 69 proteins
Krueger et al. [37]	Define tyrosine-ppome of insulin signaling pathway	SILAC, TyrP-IP, LC-MS/MS	DE, mapping to pathway	5 tp (stimulation time)	33 proteins
Matsuoka et al. [45]	DDR mediated by ATM and ATR	SILAC, phosphoSQ and phosphoTQ-IP, LC-MS/MS	DE, functional annotation, enrichment	2	905 sites, 700 proteins

(continued)

Table 2.1 (continued)

Paper	Goal	Technology	Model	Conditions	Readouts
Wu et al. [76]	Signaling downstream of NMP-ALK	IMAC, TyrP-IP + IMAC, LC-MS/MS	DE, pathway enrichment	2	4798 sites, 1548 proteins in the control; 5340 sites, 1758 proteins in the NMP-ALK cell line (506 unique to this cell line)
Jorgensen et al. [33]	Bidirectional signaling in EphB2 ephrin-B1 expressing cells	SILAC, TyrP-IP, LC-MS/MS	DE, kinase/ppase/phosphobinding proteins prediction, map to obtained network	3 cell lines	442 sites, 304 proteins

DE = differential expression (or regulation, in the case of PTMs),  
TyrP-IP = phosphotyrosine immunoprecipitation, tp = time points,  
pp = phosphoprotein, ppome = phosphoproteome,  
ppase = phosphatase, ppide = phosphopeptide

**Table 2.2** Overview of applications mentioned in this chapter using affinity based methods

Paper	Goal	Technology	Model	Conditions	Readouts
Alexopoulos et al. [3]	Difference in network topology between primary hepatocytes and hepatocellular carcinoma cell lines	xMAP	MLR	64 comb. of 7 growth factors/ck and 7 small molecule kinase inhibitors	50 ck, 17 pp
Gaudet et al. [20]; Janes et al. [29]	Link apoptotic response to signaling data, investigate aspects of data requirements	Kinase activity assays, quantitative immunoblotting, antibody microarrays, flow cytometry	PLSR	10 comb. of 3 stimuli, 13 tp	19 signaling proteins + 4 apoptotic markers
Ciaccio et al. [11]	Signaling downstream of EGF stimulation	microwestern arrays	ARACNe (correlation + information theory) and Bayesian network inference	5 EGF concentrations, 6 tp	91 phosphosites, 67 proteins
Santos et al. [63]	Architecture of MAPK network in response to NGF or EGF stimulation	Quantitative western blots	Modular-response analysis (sensitivity analysis to small perturbations)	3 siRNAs, 2 stimuli, 2 tp	3 proteins
Nelander et al. [51]	Test method on EGFR/MAPK and PI3K/AKT pathways in a breast cancer cell line	Western blots	MIMO	21 comb. of EGF + 6 inhibitors	3 proteins
Sachs et al. [59]	Test method on intracellular signaling networks of human primary naive CD4+ T cells, downstream of CD3, CD28, and LFA-1 activation	Multicolor flow cytometry (single cell)	Bayesian network inference	9 stimulators/inhibitors	11 pp and phospholipids

(continued)

Table 2.2 (continued)

Paper	Goal	Technology	Model	Conditions	Readouts
Chen et al. [9]	Quantify signal flow through ErbB-activated pathways	Multiple kinetic parameters from literature, quantitative immunoblotting, xMAP	ODEs	10 tp, 2 ligands, 3 cell lines	3 proteins
Saez-Rodriguez et al. [61]; Morris et al. [48]	Test method, understand cooperative/antagonistic interactions among ligands in hepatocellular carcinoma cells		Boolean logic [61]; constrained fuzzy logic [48]	64 comb. of 7 ck and 7 small molecule kinase inhibitors	16 proteins

tp = time points, ck = cytokine,  
pp = phosphoprotein, comb. = combinations



reporting raw data involves finding an appropriate way to store thousands of spectra (and accompanying LC retention times, and metadata) for each experiment [53]. The workflow from experiments to models can encompass multiple steps and a number of tools are available to develop data processing pipelines while maintaining the consistency of the workflow and keeping data provenance [60, 62], allowing connection with multiple modeling methods. Equally important is the development of and compliancy to standards for capturing, representing, annotating, and reporting the data and models. This should facilitate effective quality assessment, promote transparency, and enhance accessibility [69].

“Descriptive” modeling approaches mainly rely on methods from statistics and machine learning, and include for example differential expression analysis usually followed by clustering or mapping onto known networks. “Predictive” models are capable of providing estimates of the behavior of a system under a set of conditions that were not used to build the model. A very simple way to do so is using regression approaches such as PLSR, which links linearly correlated variables but do not provide mechanistic information, and can only capture linear phenomena. More detailed models can be built that include mechanistic and/or causal relationships between elements of the system that can be represented by a graph (“wiring diagram”), such as differential equations, logic-based, rule-based, or Bayesian network models. Models built upon proteomics shotgun MS data sets that have been reported so far generally belong to the descriptive category. In addition to context specific knowledge, large scale phospho-proteomics analysis by MS have generated valuable insights into the dynamics and characteristics of phosphorylation networks in signaling, which are opening new avenues for investigation.

Affinity based approaches on the other hand have produced data that have allowed extensive modeling of various (mainly well studied) systems. Although biochemical descriptions based on differential equations can provide a detailed and accurate description of signal transduction, they suffer from limitations when handling large systems, in particular due to the combinatorial complexity arising from signaling systems. Coarse graining of the system and simplifying assumptions provide ways around this but ODE/PDE models are still limited to systems of a couple of dozens of nodes. Rule-based models handle the combinatorial complexity by defining sets of rules that apply on biomolecular patterns without having to account for the full context of those patterns. This has the advantage of representing mechanistic knowledge (and assumptions) in an intuitive and explicit way, and allowing heterogeneous types of information to be incorporated into the model. However, rule-based models can only be applied to well studied systems because they rely entirely on an accumulated knowledge.

Methods that represent the system with lower level of details can provide alternatives to model bigger, not-so-well-known systems. Logic-based approaches for example represent only logical relationships between nodes in a network, and are therefore conceptually simple, computationally cheap, and causally correct [74]. Boolean logic models are limited to on/off representations of systems, but extensions such as fuzzy logic overcome this problem, albeit at the cost of increased complexity and therefore limitations in the size of the system that can be

interrogated. Finally, Bayesian networks provide a strongly statistically grounded alternative to infer signaling networks when little information about the system is available (although various levels of previous knowledge can be incorporated in the inference process). Bayesian networks can handle noise and stochasticity in the data in a natural way, but require rich data sets, which has limited their application so far to relatively small systems.

Whatever the biological question, it is very important to ask oneself the following questions before building a model: what is the scope and level of detail that I can and should model in order to (1) account for the limitations of my dataset and (2) reach the goal of my analysis. An adequate solution relies on choosing a formalism with the right level of detail to answer our question, and which yields the most interpretable results for the problem under investigation [47]. When interpreting the results of a model, it is also very important to be aware that a model is only as true as its assumptions, and that every methodology has limitations inherent to the way that they build, represent, and simulate the system.

## References

1. Aldridge BB, Burke JM, Lauffenburger DA, Sorger PK (2006) Physicochemical modelling of cell signalling pathways. *Nat Cell Biol* 8(11):1195–1203, DOI 10.1038/ncb1497, URL <http://dx.doi.org/10.1038/ncb1497>
2. Alexopoulos LG, Saez-Rodriguez J, Espelin CW (2009) High-throughput protein-based technologies and computational models for drug development, efficacy, and toxicity. John Wiley and Sons, Inc., New Jersey, pp 29–52. DOI 10.1002/9780470431818.ch2, URL <http://dx.doi.org/10.1002/9780470431818.ch2>
3. Alexopoulos LG, Saez-Rodriguez J, Cosgrove BD, Lauffenburger DA, Sorger PK (2010) Networks inferred from biochemical data reveal profound differences in toll-like receptor and inflammatory signaling between normal and transformed hepatocytes. *Mol Cell Proteom MCP* 9(9):1849–1865, DOI 10.1074/mcp.M110.000406, URL <http://www.ncbi.nlm.nih.gov/pubmed/20460255>, PMID: 20460255
4. Bansal M, Belcastro V, Ambesi-Impiombato A, di Bernardo D (2007) How to infer gene networks from expression profiles. *Mol Syst Biol* 3:78, DOI 10.1038/msb4100120, URL <http://www.ncbi.nlm.nih.gov/pubmed/17299415>, PMID: 17299415
5. Birtwistle MR, Kholodenko BN (2009) Endocytosis and signalling: a meeting with mathematics. *Mol Oncol* 3(4):308–320, DOI 10.1016/j.molonc.2009.05.009, URL <http://www.ncbi.nlm.nih.gov/pubmed/19596615>, PMID: 19596615
6. Bodenmiller B, Wanka S, Kraft C, Urban J, Campbell D, Pedrioli PG, Gerrits B, Picotti P, Lam H, Vitek O, Brusniak M, Roschitzki B, Zhang C, Shokat KM, Schlapbach R, Colman-Lerner A, Nolan GP, Nesvizhskii AI, Peter M, Loewith R, von Mering C, Aebersold R (2010) Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. *Sci Signal* 3(153):rs4, DOI 10.1126/scisignal.2001182, URL <http://www.ncbi.nlm.nih.gov/pubmed/21177495>, PMID: 21177495
7. Borisov NM, Markevich NI, Hoek JB, Kholodenko BN (2005) Signaling through receptors and scaffolds: independent interactions reduce combinatorial complexity. *Biophys J* 89(2):951–966, DOI 10.1529/biophysj.105.060533, URL <http://www.ncbi.nlm.nih.gov/pubmed/15923229>, PMID: 15923229

8. Brosch M, Choudhary J (2010) Scoring and validation of tandem MS peptide identification methods. *Meth Mol Biol* (Clifton, NJ) 604:43–53, DOI 10.1007/978-1-60761-444-9\_4, URL <http://www.ncbi.nlm.nih.gov/pubmed/20013363>, PMID: 20013363
9. Chen WW, Schoeberl B, Jasper PJ, Niepel M, Nielsen UB, Lauffenburger DA, Sorger PK (2009) Input–output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. *Mol Syst Biol* 5:239, DOI 10.1038/msb.2008.74, URL <http://www.ncbi.nlm.nih.gov/pubmed/19156131>, PMID: 19156131
10. Choudhary C, Mann M (2010) Decoding signalling networks by mass spectrometry-based proteomics. *Nat Rev Mol Cell Biol* 11(6):427–439, DOI 10.1038/nrm2900, URL <http://dx.doi.org/10.1038/nrm2900>
11. Ciaccio MF, Wagner JP, Chuu C, Lauffenburger DA, Jones RB (2010) Systems analysis of EGF receptor signaling dynamics with microwestern arrays. *Nat Meth* 7(2):148–155, DOI 10.1038/nmeth.1418, URL <http://dx.doi.org/10.1038/nmeth.1418>
12. Conzelmann H, Saez-Rodriguez J, Sauter T, Kholodenko BN, Gilles ED (2006) A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks. *BMC Bioinformatics* 7:34, DOI 10.1186/1471-2105-7-34, URL <http://www.ncbi.nlm.nih.gov/pubmed/16430778>, PMID: 16430778
13. Cutillas P, Jorgensen C (2011) Biological signalling activity measurements using mass spectrometry. *Biochem J* 434(2):189–199, DOI 10.1042/BJ20101974, URL <http://www.biochemj.org/bj/434/bj4340189.htm>
14. Danos V, Feret J, Fontana W, Harmer R, Krivine J, Biosystems P, Suprieure EN, Polytechnique E (2007) Rule-based modelling of cellular signalling. *Proc of the 18th Int Conf on Concurrency Theory (CONCUR07)*, Lecture Notes in Computer Science 4703:17–41, URL <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.107.228>
15. Diella F, Cameron S, Gemund C, Linding R, Via A, Kuster B, Sicheritz-Ponten T, Blom N, Gibson T (2004) Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. *BMC Bioinformatics* 5(1):79, DOI 10.1186/1471-2105-5-79, URL <http://www.biomedcentral.com/1471-2105/5/79>
16. Duvenaud D, Eaton D, Murphy K, Schmidt M (2009) Causal learning without DAGs. *JMLR J Mach Learn Res* URL <http://jmlr.csail.mit.edu/proceedings/papers/v6/duvenaud10a/duvenaud10a.pdf>
17. Faeder JR, Blinov ML, Hlavacek WS (2009) Rule-based modeling of biochemical systems with BioNetGen. *Meth Mol Biol* (Clifton, NJ) 500:113–167, DOI 10.1007/978-1-59745-525-1\_5, URL <http://www.ncbi.nlm.nih.gov/pubmed/19399430>, PMID: 19399430
18. Feret J, Danos V, Krivine J, Harmer R, Fontana W (2009) Internal coarse-graining of molecular systems. *Proc Natl Acad Sci USA* 106(16):6453–6458, DOI 10.1073/pnas.0809908106, URL <http://www.ncbi.nlm.nih.gov/pubmed/19346467>, PMID: 19346467
19. Gat-Viks I, Shamir R (2007) Refinement and expansion of signaling pathways: the osmotic response network in yeast. *Genome Res* 17(3):358–367, DOI 10.1101/gr.5750507, URL <http://www.ncbi.nlm.nih.gov/pubmed/17267811>, PMID: 17267811
20. 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 Proteom MCP* 4(10):1569–1590, DOI 10.1074/mcp.M500158-MCP200, URL <http://www.ncbi.nlm.nih.gov/pubmed/16030008>, PMID: 16030008
21. Gnad F, Ren S, Cox J, Olsen JV, Macek B, Oroschi M, Mann M (2007) PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol* 8(11):R250, DOI 10.1186/gb-2007-8-11-r250, URL <http://www.ncbi.nlm.nih.gov/pubmed/18039369>, PMID: 18039369
22. Gouw JW, Krijgsveld J, Heck AJR (2010) Quantitative proteomics by metabolic labeling of model organisms. *Mol Cell Proteom MCP* 9(1):11–24, DOI 10.1074/mcp.R900001-MCP200, URL <http://www.ncbi.nlm.nih.gov/pubmed/19955089>, PMID: 19955089
23. Gstaiger M, Aebersold R (2009) Applying mass spectrometry-based proteomics to genetics, genomics and network biology. *Nat Rev Genet* 10(9):617–627, DOI 10.1038/nrg2633, URL <http://dx.doi.org/10.1038/nrg2633>

24. Harrison C (2008) High-content screening: integrating information. *Nat Rev Drug Discov* 7(2):121, DOI 10.1038/nrd2522, URL <http://dx.doi.org/10.1038/nrd2522>
25. Hlavacek WS, Faeder JR, Blinov ML, Posner RG, Hucka M, Fontana W (2006) Rules for modeling signal-transduction systems. *Science's STKE: Signal Transduct Knowl Environ* 2006(344):ref6, DOI 10.1126/stke.3442006ref6, URL <http://www.ncbi.nlm.nih.gov/pubmed/16849649>, PMID: 16849649
26. Hornbeck PV, Chabra I, Kornhauser JM, Skrzypek E, Zhang B (2004) PhosphoSite: a bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* 4(6):1551–1561, DOI 10.1002/pmic.200300772, URL <http://www.ncbi.nlm.nih.gov/pubmed/15174125>, PMID: 15174125
27. Huang PH, Mukasa A, Bonavia R, Flynn RA, Brewer ZE, Cavenee WK, Furnari FB, White FM (2007) Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma. *Proc Natl Acad Sci USA* 104(31):12,867–12,872, DOI 10.1073/pnas.0705158104, URL <http://www.ncbi.nlm.nih.gov/pubmed/17646646>, PMID: 17646646
28. Hyduke DR, Palsson B (2010) Towards genome-scale signalling-network reconstructions. *Nat Rev Genet* 11(4):297–307, DOI 10.1038/nrg2750, URL <http://dx.doi.org/10.1038/nrg2750>
29. 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* (New York, NY) 310(5754):1646–1653, DOI 10.1126/science.1116598, URL <http://www.ncbi.nlm.nih.gov/pubmed/16339439>, PMID: 16339439
30. Jensen ON (2006) Interpreting the protein language using proteomics. *Nat Rev Mol Cell Biol* 7(6):391–403, DOI 10.1038/nrm1939, URL <http://www.ncbi.nlm.nih.gov/pubmed/16723975>, PMID: 16723975
31. de Jong H (2002) Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol: J Comput Mol Cell Biol* 9(1):67–103, DOI 10.1089/10665270252833208, URL <http://www.ncbi.nlm.nih.gov/pubmed/11911796>, PMID: 11911796
32. Jorgensen C, Linding R (2010) Simplistic pathways or complex networks? *Curr Opin Genet Dev* 20(1):15–22, DOI 10.1016/j.gde.2009.12.003, URL <http://www.ncbi.nlm.nih.gov/pubmed/20096559>, PMID: 20096559
33. Jorgensen C, Sherman A, Chen GI, Pasculescu A, Poliakov A, Hsiung M, Larsen B, Wilkinson DG, Linding R, Pawson T (2009) Cell-specific information processing in segregating populations of eph receptor ephrin-expressing cells. *Science* (New York, NY) 326(5959):1502–1509, DOI 10.1126/science.1176615, URL <http://www.ncbi.nlm.nih.gov/pubmed/20007894>, PMID: 20007894
34. Joughin BA, Cheung E, Karuturi RKM, Saez-Rodriguez J, Lauffenburger DA, Liu ET (2010) Cellular regulatory networks, systems biomedicine – Chapter 4. Academic Press, San Diego, pp 57–108, DOI 10.1016/B978-0-12-372550-9.00004-3, URL <http://www.sciencedirect.com/science/article/pii/B9780123725509000043>
35. Kauffman S (1969) Homeostasis and differentiation in random genetic control networks. *Nature* 224(5215):177–178, URL <http://www.ncbi.nlm.nih.gov/pubmed/5343519>, PMID: 5343519
36. Kholodenko BN (2006) Cell-signalling dynamics in time and space. *Nat Rev Mol Cell Biol* 7(3):165–176, DOI 10.1038/nrm1838, URL <http://www.ncbi.nlm.nih.gov/pubmed/16482094>, PMID: 16482094
37. Krueger M, Kratchmarova I, Blagoev B, Tseng Y, Kahn CR, Mann M (2008) Dissection of the insulin signaling pathway via quantitative phosphoproteomics. *Proc Natl Acad Sci USA* 105(7):2451–2456, DOI 10.1073/pnas.0711713105, URL <http://www.ncbi.nlm.nih.gov/pubmed/18268350>, PMID: 18268350
38. Krutzik PO, Clutter MR, Trejo A, Nolan GP (2011) Fluorescent cell barcoding for multiplex flow cytometry. In: Robinson JP, Darzynkiewicz Z, Dobrucki J, Hyun WC, Nolan JP, Orfao A, Rabinovitch PS (eds) *Current protocols in cytometry*. John Wiley & Sons, Inc., Hoboken, NJ, USA, URL <http://www.currentprotocols.com/protocol/cy0631>

39. Li P, Zhang C, Perkins EJ, Gong P, Deng Y (2007) Comparison of probabilistic boolean network and dynamic bayesian network approaches for inferring gene regulatory networks. *BMC Bioinformatics* 8 Suppl 7:S13, DOI 10.1186/1471-2105-8-S7-S13, URL <http://www.ncbi.nlm.nih.gov/pubmed/18047712>, PMID: 18047712
40. Lindling R, Jensen LJ, Ostheimer GJ, van Vugt MATM, Jorgensen C, Miron IM, Diella F, Colwill K, Taylor L, Elder K, Metalnikov P, Nguyen V, Pasculescu A, Jin J, Park JG, Samson LD, Woodgett JR, Russell RB, Bork P, Yaffe MB, Pawson T (2007) Systematic discovery of in vivo phosphorylation networks. *Cell* 129(7):1415–1426, DOI 10.1016/j.cell.2007.05.052, URL <http://www.ncbi.nlm.nih.gov/pubmed/17570479>, PMID: 17570479
41. Malmström J, Lee H, Aebersold R (2007) Advances in proteomic workflows for systems biology. *Curr Opin Biotechnol* 18(4):378–384, DOI 10.1016/j.copbio.2007.07.005, URL <http://www.ncbi.nlm.nih.gov/pubmed/17698335>, PMID: 17698335
42. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Favera RD, Califano A (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7 Suppl 1:S7, DOI 10.1186/1471-2105-7-S1-S7, URL <http://www.ncbi.nlm.nih.gov/pubmed/16723010>, PMID: 16723010
43. Markowetz F (2010) How to understand the cell by breaking it: network analysis of gene perturbation screens. *PLoS Comput Biol* 6(2):e1000655, DOI 10.1371/journal.pcbi.1000655, URL <http://www.ncbi.nlm.nih.gov/pubmed/20195495>, PMID: 20195495
44. Markowetz F, Spang R (2007) Inferring cellular networks – A review. *BMC Bioinformatics* 8 Suppl 6:S5, DOI 10.1186/1471-2105-8-S6-S5, URL <http://www.ncbi.nlm.nih.gov/pubmed/17903286>, PMID: 17903286
45. Matsuoka S, Ballif BA, Smogorzewska A, McDonald ER, Hurov KE, Luo J, Bakalarski CE, Zhao Z, Solimini N, Lerenthal Y, Shiloh Y, Gygi SP, Elledge SJ (2007) ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science (New York, NY)* 316(5828):1160–1166, DOI 10.1126/science.1140321, URL <http://www.ncbi.nlm.nih.gov/pubmed/17525332>, PMID: 17525332
46. Miller ML, Jensen LJ, Diella F, Jorgensen C, Tinti M, Li L, Hsiung M, Parker SA, Bordeaux J, Sicheritz-Ponten T, Olhovsky M, Pasculescu A, Alexander J, Knapp S, Blom N, Bork P, Li S, Cesareni G, Pawson T, Turk BE, Yaffe MB, Brunak S, Lindling R (2008) Linear motif atlas for phosphorylation-dependent signaling. *Sci Signal* 1(35):ra2, DOI 10.1126/scisignal.1159433, URL <http://www.ncbi.nlm.nih.gov/pubmed/18765831>, PMID: 18765831
47. Morris MK, Saez-Rodriguez J, Sorger PK, Lauffenburger DA (2010) Logic-based models for the analysis of cell signaling networks. *Biochemistry* 49(15):3216–3224, DOI 10.1021/bi902202q, URL <http://www.ncbi.nlm.nih.gov/pubmed/20225868>, PMID: 20225868
48. Morris MK, Saez-Rodriguez J, Clarke DC, Sorger PK, Lauffenburger DA (2011) Training signaling pathway maps to biochemical data with constrained fuzzy logic: quantitative analysis of liver cell responses to inflammatory stimuli. *PLoS Comput Biol* 7(3):e1001099, DOI 10.1371/journal.pcbi.1001099, URL <http://dx.doi.org/10.1371/journal.pcbi.1001099>
49. Mukherjee S, Speed TP (2008) Network inference using informative priors. *Proc Natl Acad Sci* 105(38):14,313–14,318, DOI 10.1073/pnas.0802272105, URL <http://www.pnas.org/content/105/38/14313.abstract>
50. Naegle KM, Gymrek M, Joughin BA, Wagner JP, Welsch RE, Yaffe MB, Lauffenburger DA, White FM (2010) PTMScout, a web resource for analysis of high throughput post-translational proteomics studies. *Mol Cell Proteom MCP* 9(11):2558–2570, DOI 10.1074/mcp.M110.001206, URL <http://www.ncbi.nlm.nih.gov/pubmed/20631208>, PMID: 20631208
51. Nelander S, Wang W, Nilsson B, She Q, Pratilas C, Rosen N, Gennemark P, Sander C (2008) Models from experiments: combinatorial drug perturbations of cancer cells. *Mol Syst Biol* 4:216, DOI 10.1038/msb.2008.53, URL <http://www.ncbi.nlm.nih.gov/pubmed/18766176>, PMID: 18766176
52. Nesvizhskii AI, Aebersold R (2005) Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteom MCP* 4(10):1419–1440, DOI 10.1074/mcp.R500012-MCP200, URL <http://www.ncbi.nlm.nih.gov/pubmed/16009968>, PMID: 16009968

53. Olsen JV, Mann M (2011) Effective representation and storage of mass Spectrometry-Based proteomic data sets for the scientific community. *Sci Signal* 4(160):pe7, DOI 10.1126/scisignal.2001839, URL <http://stke.sciencemag.org/cgi/content/abstract/sigtrans;4/160/pe7>
54. Olsen JV, Blagoev B, Gnäd F, Macek B, Kumar C, Mortensen P, Mann M (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127(3):635–648, DOI 10.1016/j.cell.2006.09.026, URL <http://www.ncbi.nlm.nih.gov/pubmed/17081983>, PMID: 17081983
55. Ong S, 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 Proteom* 1(5):376–386, DOI 10.1074/mcp.M200025-MCP200, URL <http://www.mcponline.org/content/1/5/376.abstract>
56. Picotti P, Bodenmiller B, Mueller LN, Domon B, Aebersold R (2009) Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* 138(4):795–806, DOI 10.1016/j.cell.2009.05.051, URL <http://www.ncbi.nlm.nih.gov/pubmed/19664813>, PMID: 19664813
57. Prill RJ, Marbach D, Saez-Rodriguez J, Sorger PK, Alexopoulos LG, Xue X, Clarke ND, Altan-Bonnet G, Stolovitzky G (2010) Towards a rigorous assessment of systems biology models: the DREAM3 challenges. *PloS One* 5(2):e9202, DOI 10.1371/journal.pone.0009202, URL <http://www.ncbi.nlm.nih.gov/pubmed/20186320>, PMID: 20186320
58. Rosario AMD, White FM (2010) Quantifying oncogenic phosphotyrosine signaling networks through systems biology. *Curr Opin Genet Dev* 20(1):23–30, DOI 10.1016/j.gde.2009.12.005, URL <http://www.ncbi.nlm.nih.gov/pubmed/20074929>, PMID: 20074929
59. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP (2005) Causal Protein-Signaling networks derived from multiparameter Single-Cell data. *Science* 308(5721):523–529, DOI 10.1126/science.1105809, URL <http://www.sciencemag.org/content/308/5721/523.abstract>
60. Saez-Rodriguez J, Goldsipe A, Muhlich J, Alexopoulos LG, Millard B, Lauffenburger DA, Sorger PK (2008) Flexible informatics for linking experimental data to mathematical models via DataRail. *Bioinformatics* 24(6):840–847, DOI 10.1093/bioinformatics/btn018, URL <http://bioinformatics.oxfordjournals.org/content/24/6/840.abstract>
61. Saez-Rodriguez J, Alexopoulos LG, Epperlein J, Samaga R, Lauffenburger DA, Klamt S, Sorger PK (2009) Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol Syst Biol* 5:331, DOI 10.1038/msb.2009.87, URL <http://www.ncbi.nlm.nih.gov/pubmed/19953085>, PMID: 19953085
62. Saez-Rodriguez J, Alexopoulos LG, Stolovitzky G (2011) Setting the standards for signal transduction research. *Sci Signal* 4(160):pe10, DOI 10.1126/scisignal.2001844, URL <http://stke.sciencemag.org/cgi/content/abstract/sigtrans;4/160/pe10>
63. Santos SDM, Verveer PJ, Bastiaens PIH (2007) Growth factor-induced MAPK network topology shapes erk response determining PC-12 cell fate. *Nat Cell Biol* 9(3):324–330, DOI 10.1038/ncb1543, URL <http://www.ncbi.nlm.nih.gov/pubmed/17310240>, PMID: 17310240
64. Smith RD, Anderson GA, Lipton MS, Pasa-Tolic L, Shen Y, Conrads TP, Veenstra TD, Udseth HR (2002) An accurate mass tag strategy for quantitative and high-throughput proteome measurements. *Proteomics* 2(5):513–523, DOI 10.1002/1615-9861(200205)2:5<513::AID-PROT513>3.0.CO;2-W, URL <http://www.ncbi.nlm.nih.gov/pubmed/11987125>, PMID: 11987125
65. Sneddon MW, Faeder JR, Emonet T (2011) Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. *Nat Meth* 8(2):177–183, DOI 10.1038/nmeth.1546, URL <http://www.ncbi.nlm.nih.gov/pubmed/21186362>, PMID: 21186362
66. Steen H, Mann M (2004) The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* 5(9):699–711, DOI 10.1038/nrm1468, URL <http://www.ncbi.nlm.nih.gov/pubmed/15340378>, PMID: 15340378
67. Tan CSH, Jrgensen C, Lindling R (2010) Roles of “junk phosphorylation” in modulating biomolecular association of phosphorylated proteins? *Cell Cycle* (Georgetown, Tex) 9(7):1276–1280, URL <http://www.ncbi.nlm.nih.gov/pubmed/20234177>, PMID: 20234177



68. Tanner SD, Ornatsky O, Bandura DR, Baranov VI (2007) Multiplex bio-assay with inductively coupled plasma mass spectrometry: towards a massively multivariate single-cell technology. *Spectrochim Acta B* 62(3):188–195, DOI 10.1016/j.sab.2007.01.008, URL <http://www.sciencedirect.com/science/article/B6THN-4N0HJDH-1/2/05bbd4e8f7b003df4d258be40015b7ba>
69. Taylor CF, Field D, Sansone S, Aerts J, Apweiler R, Ashburner M, Ball CA, Binz P, Bogue M, Booth T, Brazma A, Brinkman RR, Clark AM, Deutsch EW, Fiehn O, Fostel J, Ghazal P, Gibson F, Gray T, Grimes G, Hancock JM, Hardy NW, Hermjakob H, Julian RK, Kane M, Kettner C, Kinsinger C, Kolker E, Kuiper M, Novere NL, Leebens-Mack J, Lewis SE, Lord P, Mallon A, Marthandan N, Masuya H, McNally R, Mehrle A, Morrison N, Orchard S, Quackenbush J, Reecy JM, Robertson DG, Rocca-Serra P, Rodriguez H, Rosenfelder H, Santoyo-Lopez J, Scheuermann RH, Schober D, Smith B, Snape J, Stoeckert CJ, Tipton K, Sterk P, Untergasser A, Vandesompele J, Wiemann S (2008) Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Nat Biotechnol* 26(8):889–896, DOI 10.1038/nbt.1411, URL <http://dx.doi.org/10.1038/nbt.1411>
70. Thompson A, Schfer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Hamon C (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* 75(8):1895–1904, DOI 10.1021/ac0262560, URL <http://dx.doi.org/10.1021/ac0262560>
71. Vignali DA (2000) Multiplexed particle-based flow cytometric assays. *J Immunol Meth* 243(1–2):243–255, URL <http://www.ncbi.nlm.nih.gov/pubmed/10986418>, PMID: 10986418
72. Vizcano JA, Ct R, Reisinger F, Foster JM, Mueller M, Rameseder J, Hermjakob H, Martens L (2009) A guide to the proteomics identifications database proteomics data repository. *Proteomics* 9(18):4276–4283, DOI 10.1002/pmic.200900402, URL <http://www.ncbi.nlm.nih.gov/pubmed/19662629>, PMID: 19662629
73. Vogel C, de Sousa Abreu R, Ko D, Le S, Shapiro BA, Burns SC, Sandhu D, Boutz DR, Marcotte EM, Penalva LO (2010) Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Mol Syst Biol* 6, DOI 10.1038/msb.2010.59, URL <http://dx.doi.org/10.1038/msb.2010.59>
74. Watterson S, Marshall S, Ghazal P (2008) Logic models of pathway biology. *Drug Discov Today* 13(9–10):447–456, DOI 10.1016/j.drudis.2008.03.019, URL <http://www.ncbi.nlm.nih.gov/pubmed/18468563>, PMID: 18468563
75. Wolf-Yadlin A, Sevecka M, MacBeath G (2009) Dissecting protein function and signaling using protein microarrays. *Curr Opin Chem Biol* 13(4):398–405, DOI 10.1016/j.cbpa.2009.06.027, URL <http://www.ncbi.nlm.nih.gov/pubmed/19660979>, PMID: 19660979
76. Wu F, Wang P, Zhang J, Young LC, Lai R, Li L (2010) Studies of phosphoproteomic changes induced by nucleophosmin-anaplastic lymphoma kinase (ALK) highlight deregulation of tumor necrosis factor (TNF)/Fas/TNF-related apoptosis-induced ligand signaling pathway in ALK-positive anaplastic large cell lymphoma. *Mol Cell Proteom MCP* 9(7):1616–1632, DOI 10.1074/mcp.M000153-MCP201, URL <http://www.ncbi.nlm.nih.gov/pubmed/20393185>, PMID: 20393185



<http://www.springer.com/978-1-4419-7209-5>

Advances in Systems Biology

Goryanin, I.I.; Goryachev, A. (Eds.)

2012, XXII, 682 p., Hardcover

ISBN: 978-1-4419-7209-5