Invited Review

Leveraging systems biology approaches in clinical pharmacology

Ioannis N. Melas[a,b]**, Kosmas Kretsos**[c]**, and Leonidas G. Alexopoulos**[a,b,*]
[a]*National Technical University of Athens, Athens, Greece*
[b]*Protatonce Ltd, Athens, Greece*
[c]*UCB Pharma, Slough, UK*

ABSTRACT: Computational modeling has been adopted in all aspects of drug research and development, from the early phases of target identification and drug discovery to the late-stage clinical trials. The different questions addressed during each stage of drug R&D has led to the emergence of different modeling methodologies. In the research phase, systems biology couples experimental data with elaborate computational modeling techniques to capture lifecycle and effector cellular functions (e.g. metabolism, signaling, transcription regulation, protein synthesis and interaction) and integrates them in quantitative models. These models are subsequently used in various ways, i.e. to identify new targets, generate testable hypotheses, gain insights on the drug's mode of action (MOA), translate preclinical findings, and assess the potential of clinical drug efficacy and toxicity. In the development phase, pharmacokinetic/pharmacodynamic (PK/PD) modeling is the established way to determine safe and efficacious doses for testing at increasingly larger, and more pertinent to the target indication, cohorts of subjects. First, the relationship between drug input and its concentration in plasma is established. Second, the relationship between this concentration and desired or undesired PD responses is ascertained. Recognizing that the interface of systems biology with PK/PD will facilitate drug development, systems pharmacology came into existence, combining methods from PK/PD modeling and systems engineering explicitly to account for the implicated mechanisms of the target system in the study of drug–target interactions. Herein, a number of popular system biology methodologies are discussed, which could be leveraged within a systems pharmacology framework to address major issues in drug development. © 2013 The Authors. *Biopharmaceutics & Drug Disposition* published by John Wiley & Sons, Ltd.

Key words: systems pharmacology; systems biology; modeling of signaling pathways; identification of drug mode of action; prediction of clinical drug efficacy

Introduction

Computational modeling has been adopted in all aspects of drug research and development, from the early stages of target identification and drug discovery, up to phase II–IV of clinical trials. In the early stages, computational modeling in the form of systems biology, fueled by recent advances in 'omics' technologies, constructs predictive models to integrate major cellular functions and monitor how these are altered in disease; identifies new drug targets; predicts the potential of clinical efficacy and toxicity of uncharacterized compounds and probes their mode of action (MOA). Systems biology employs elaborate methodologies to exploit and integrate prior knowledge of the

interrogated system (disease, cell type, a specific tissue, etc. and their interactions with compounds of interest) and extracts qualitative or quantitative features that will lead to its better understanding, eventually facilitating drug development. The preclinical and clinical development, on the other hand, follows a completely different paradigm. Scientists, clinical pharmacologists and physicians have to decide on issues such as: What is the optimal drug candidate to advance in the next phase of clinical trials? What dose/regimen will maximize efficacy and minimize toxicity? What are the right patients to treat with the drug? Because of the nature and consequences of these questions and because of the existing staged paradigm of drug development [1], PK/PD modeling, an empirical data-driven approach based on cardinal pharmacological principles [2], has gained wide acceptance: PK/PD modeling first establishes the relationship between drug input and its concentration in plasma and second, establishes the relationship between this drug plasma concentration and a desired or undesired response, be it a biomarker, a clinical endpoint or an adverse event, e.g. tumor growth or cell apoptosis.

Traditionally, systems biology and PK/PD modeling existed in 'parallel universes' [3]. However, it is becoming evident that their role is complementary and that a lot could be gained by their integration. For example, PK/PD overlooks pertinent biological network aspects such as the molecular basis of the disease, the way in which cellular processes are orchestrated by intricate signaling mechanisms, and how these mechanisms cross-react. This severely limits the general applicability of the model and disallows extrapolations and wider pharmacological and biological insights. On the other hand, PK/PD is a relatively tractable and pharmacologically sound method that manages to tame and quantify the uncertainty around the point estimates of the model parameters as well as the inter-individual/biological and unexplained variability inherent in the data. Thus, it can quantify the dose–exposure–response relationship and assess the robustness of the model. The latter is not always true in systems biology. Systems biology figures out the qualitative and quantitative aspects of the molecular mechanisms of disease by leveraging extensive and diverse data from *in vitro*

experiments, but that imposes limitations, especially when there is limited understanding of how these data are incorporated in the physiology of the target organism and how they affect its clinical response. Thus, depending on the system under study, systems biology approaches may lead to intractable or ill-defined models with low confidence on their parameters making their use for robust quantitative predictions of clinical response risky. Systems pharmacology emerged to form this exact interface between PK/PD and systems biology [4,5]. It combines methods from PK/PD modeling and systems engineering explicitly to account for the implicated mechanisms of the target system in the study of drug–target interactions, albeit in a tractable, robust way. In more detail, instead of ignoring the biology of the target organism and focusing only on the data-driven correlation of drug exposure and clinical response, systems pharmacology employs mechanistic methods to capture key properties of the system (e.g. focused biology around the target or biomarker), increasing the applicability and relevance of the model.

Herein, we first present the current PK/PD methodologies applied in drug development and then discuss a number of systems biology approaches that could be leveraged within a systems pharmacology framework to facilitate: (i) the modeling of signaling pathways, (ii) identification of drug MOA and (iii) prediction of clinical drug efficacy and toxicity.

Computational Modeling in the Clinic

In the clinical phase of drug development, computational modeling takes place mostly in the form of PK/PD. This is now a well-defined discipline explained in established textbooks [2,6], thus, the technical details will not be covered here. Physiologically based pharmacokinetic (PBPK) modeling, an approach lying between standard PK modeling and systems pharmacology, warrants a brief discussion. PBPK uses compartments that correspond to specific tissues/organs and models drug distribution between them in a physiologically realistic manner using the cardiovascular system [7,8]. Moreover, recent advancements in the predictability of key pharmacokinetic

parameters from human *in vitro* data and in the availability of dedicated software platforms and associated databases, allowed PBPK to construct even more detailed models, predicting the time dependent plasma concentration of the drug, drug–drug interactions and the effects of age, genetics and disease to the kinetics of the drug.

Recently, a few mechanistic models have been proposed that capture the processes governing the transduction of target activation into the response *in vivo* [9]. These models employ concepts from dynamic systems analysis (such as ordinary differential equations (ODE)s modeling) to model signaling cascades or even homeostatic feedback (transduction models).

It has become clear in recent years that PK/PD models are evolving to account for the mechanisms of the target organism [4]. Systems pharmacology works in this direction, attempting to incorporate methodologies from systems engineering and systems biology with PK/PD modeling to widen its applicability and to facilitate the prediction of drug action [4,5]. Systems pharmacology in the preclinical phase could aid in the design of clinical companion tests to uncover sensitivity (or resistance) to different drugs; or screen high-risk drugs out of the development pipeline as early as possible, de-risking programs ahead of phase II and saving valuable resources. Unfortunately, the clinical applicability of systems approaches is constrained by the limited sample availability. In contrast to the discovery phase where extensive *in vitro* experiments can be carried out, during clinical development computational models are forced to work with a few biomarkers expressed in blood (e.g. cytokine releases in the bloodstream) and only occasionally with biopsy samples. Moreover, biomarkers in blood are typically characterized by a low signal to noise ratio, resulting in obscure predictions. A workaround to some of these limitations may involve the proteomic technologies that provide high content data with minimum sample requirements, such as the xMAP technology, protein microarrays and flow cytometry (see Box 1). Some of them have already been used in the clinic, mostly in the quest for personalized medicine. Leveraging these technologies within a systems pharmacology framework could provide the experimental data necessary for the construction of more detailed and biologically relevant models.

Box. 1. **Phosphoproteomic technologies**.

Phosphoproteomic technologies play a key role in computational modeling for drug development, since they provide high content data at the level where most of the modern drugs act [49]. As a result, all forms of mechanistic modeling for the study of signaling pathways and the identification of drug MOA, leverages phosphoproteomic data. The most commonly employed phosphoproteomic technologies in drug development, that are also suitable for the clinical phase and the analysis of biopsy samples are protein microarrays, xMAP technology and flow cytometry.

Protein microarrays [34,35] typically employ a glass slide, on top of which the capture antibody is spotted. Then, the sample bathes the slide and proteins from the sample are immobilized on the matching antibodies. At the final step of the assay, a secondary antibody (which is biotinylated) bathes the slide and binds to the corresponding proteins. The amount of proteins on the slide is measured by a plate reader, providing an estimate of protein abundance; while, if one of the two antibodies is anti-phospho, the phosphorylation of the protein is estimated instead. The use of robotic systems to spot the antibodies on the slide allows microarrays to measure up to 1000 proteins per sample, essentially making antibody availability the limiting factor of this platform. A variation of protein microarrays called Reverse Phase Protein Arrays (RPPA) employs the spotting of the sample on the glass slide; then an antibody solution, which is biotinylated, bathes the slide and binds to the matching proteins. In this manner thousands of samples can be screened at the same time but measuring only a single signal.

xMAP technology [36,64], is an antibody-based, suspension array technology. xMAP employs polystyrene beads as a substrate, on top of which capture antibodies are coupled; the beads are color-coded, so every bead color corresponds to a different capture antibody (signal). The sample is plated on 96 or 384 well plates and beads of different colors are multiplexed and suspended with the sample. Proteins from the sample bind to the capture antibody on the bead surface, while the rest of them are washed away. Then, a secondary antibody, which is biotinylated, is introduced and binds to the immobilized proteins, thus completing a sandwich assay. Then another washing step follows and the bead–protein–secondary antibody construct goes through the xMAP detection system in which two lasers are used. One excites the bead's red color and one excites the fluorophore's green color, two photomultipliers collect the emissions and provide an estimate of protein abundance for each signal. With xMAP technology, up to 30 signals may be measured in each well, providing high sample and signal throughput.

Flow cytometry [37] is a single cell technology that employs the suspension of cells, properly labeled with fluorescent chemicals, in a stream of liquid going through a detection system. The detection system uses a laser to excite the fluorophore and a photomultiplier measures the emitted signature. To measure phosphorylation activity, phospho-specific fluorescent antibodies are used that bind to the target proteins. If more than one signals are to be quantified at the same time, the fluorophores must emit in different wavelengths (colors). Using polychromatic flow cytometry, Perez *et al.* in [37] measured a total of 11 proteins (members of the MAPK family) in both artificially and physiologically perturbed peripheral blood mononuclear cells (PBMCs).

Computational Modeling in the Discovery Phase

Computational modeling in the discovery phase employs elaborate methodologies to exploit prior knowledge of the interrogated system (disease, cell type, a specific tissue, a compound of interest, etc.) and extracts qualitative features that will lead to its better understanding and ultimately facilitate drug development. The employed methodologies can be broken down into two classes: (i) the data driven methods and (ii) the mechanism driven methods. Data driven methods exploit extensive datasets and use straightforward approaches to extract interpretable features of the interrogated system. These approaches include mostly machine learning algorithms (clustering analysis, classification, Bayesian inference), regression methods, methods from information theory (mutual information) or optimization algorithms. These methods are agnostic to the underlying biology since they ignore the mechanisms that define system behavior. Instead they trust the experimental data in capturing all relevant information on the interrogated system and focus on modeling these data. On the other hand, mechanism driven methods employ a mathematical formalism to model an often elementary process of the system (e.g. signal transduction from one phosphoprotein to the next, metabolism of one substance to another, expression of a gene from a transcription factor, etc.) and then integrate all these processes in a computable model such as the signaling pathway downstream of a receptor of interest, a metabolic pathway, or a gene expression network. Typically, data driven methods are employed to construct predictive models on a higher system level (e.g. gene regulatory networks numbering tens of thousands of nodes) but not very detailed. Mechanism driven models are used to construct more detailed models, albeit around a narrow region of interest. The trade-off between relevance and tractability dictates that the more detailed the model, the narrower the region. The reasons for this discrimination is, first, that mechanism driven methods require high complexity data for the interrogated system (e.g. perturbation data, many time points etc.), implying a large number of samples, that very often come at the cost of a small number of

measured signals; and second, the parameter estimation problem that eventually has to be solved, becomes very challenging (computationally) for large models. On the other hand, data driven methods are computationally simpler and do not require such complex, multi-dimensional data, allowing the construction of models on a whole systems level. In the following paragraphs we review methodologies that address: (i) the modeling of signaling pathways, (ii) identification of drug MOA and (iii) prediction of clinical drug efficacy and toxicity; and discuss how these could be leveraged in a systems pharmacology framework to facilitate drug development in the clinic.

Modeling signal transduction pathways by leveraging experimental data

Modeling of signaling pathways refers to the process of identifying relationships that describe how signal propagates from one protein to the next, ultimately explaining the way cells respond to factors of their biochemical microenvironment [10]. The study of signaling pathways is of the utmost importance in both the discovery and the clinical phase of drug development, however, it takes place very differently in the two phases. In the discovery phase, typically *in vitro* data are used either to construct mechanistic models that describe in detail how signal transduction takes place in the cell type/tissue of interest and how this is affected by disease, or to construct less detailed but more extensive models, integrating dozens of pathways that orchestrate all major cellular processes. In the clinical phase, PD modeling of biomarkers mostly expressed in blood and occasionally in biopsy samples, aims at the identification of signaling events only at the very narrow region of the pathway where the interrogated drug is expected to act (in the neighborhood of a few predefined, accessible biomarkers). For the study of this region, mechanistic PD modeling employs detailed methods such as ODE modeling and succeeds in capturing the dynamics of the implicated reactions; it then correlates the expression of these biomarkers with clinical endpoints. For example, Ramakrishnan *et al.* [11] built an ODE model to study the pharmacodynamic effects of methylprednisolone, as a series of events initiating

at the cytosolic glucocorticoid receptor, going through the heat shock proteins to the nucleus and resulting to the enhanced expression of STAT, with STAT being the ultimate pharmacodynamics endpoint.

Pharmacodynamic modeling often fails to see the general context and broader processes that these biomarkers are a part of, and regulated by. In the following paragraphs we present methodologies for the modeling of signaling pathways applied in early stages of drug development, but also suitable for the clinical phase.

The methodologies for pathway modeling can be broken down into two classes: (i) Data inference methods, and (ii) mechanistic methods [12].

Data inference methods. typically employ principal component analysis (PCA), partial least squares regression (PLSR), clustering, self-organizing maps and network inference algorithms such as mutual information (MI) and Bayesian inference, to identify cross-talks between the signaling molecules in the pathway, as these are captured in the experimental data at hand. They do not require any form of prior knowledge of the proteins' connectivity for their implementation, but extract all their predictions based on the training dataset. In more detail, methods such as PLSR and PCA perform dimensionality reduction on the experimental dataset by projecting it to the dimensions of maximum variation. This facilitates data interpretation and the identification of qualitative trends in the signaling process [13]. PLSR also performs regression, correlating a perturbations matrix X with the signaling dataset Y, identifying in this manner the features of the perturbation matrix (stimuli, drugs etc.) that best explain the variance in the measurements [14]. Clustering and self-organizing maps employ a distance metric to identify signals in the experimental dataset that respond in similar fashion across all samples. Then a threshold is introduced above which these similarities imply an interaction between the two signals [12,15]. Mutual information instead of using distance metrics uses an integral function of the joint probability of any two signals over all samples, to calculate the dependency between them. A threshold is then introduced, in similar fashion to the clustering methods, above which the dependency of the

two signals is considered strong enough to imply an interaction between them [16]. Finally, Bayesian inference is one of the most powerful methods for data inference. It employs the Bayes rule to model the probability of an arbitrary signal to be active as a function of its upstream signals [17,18]. These probabilities in the Bayesian framework are called conditional probability distributions (CPDs) and essentially capture the proteins' connectivity in the signaling pathway. The CPDs can be learned from the data using a training algorithm such as the expectation maximization algorithm.

Mechanistic methods. implement a mathematical formalism to model how signal propagates from one protein to the next within the signaling pathway. Typically these formalisms include that of ODEs, some form of logic modeling (such as Boolean logic or constrained fuzzy logic), or employ custom rules to model the signaling pathway as an interaction graph. Either of these formalisms essentially translates the pathway from an abstract graphical representation of protein interactions (as obtained from the literature) into an executable model of the cell's signaling mechanisms, capable of simulating the signal flow from the receptor level, through the several kinases implicated in the signaling process, all the way into the transcription level. Mechanistic models offer the clear advantage of *in silico* experiments, i.e. *what will the signaling process be like if I stimulated the cells with a growth factor, or with a candidate drug?* The difference between the various formalisms lies in their perception of the signal transduction processes. The ODEs are one of the most detailed formalisms, employing the law of mass action kinetics to calculate the proteins' activation state over time [19,20]. Boolean logic assumes binary (0/1) values for the activation state of the included proteins and uses logic gates (AND/OR/NOT) to model the proteins' connectivity in the pathway. Then, signal flow is simulated by imposing boundary conditions on input nodes (receptors, or targets of compounds) and by propagating the signal downstream via the logic gates [21–25]. Constrained fuzzy logic also employs logic gates (AND/OR/NOT) but also incorporates a transfer function (typically a sigmoid curve) to calculate

the activation state of a given node as a function of the activation state of its upstream nodes [26,27]. Finally, the simplest representation of a signaling pathway is that of a graph model and in particular, that of a signed directed graph. In signed directed graphs each edge indicates either a positive or a negative effect of one node upon another. In the work by Melas *et al*. (Detecting and removing inconsistencies between experimental data and signaling network topologies using integer linear programming on interaction graphs, accepted in *PLoS Comp Biol*, 2013), a set of rules is proposed based on the definition of the signed directed graphs that models signal transduction from one node to the next, implemented as a set of linear constraints.

In principle, mechanistic methods could be applied without using experimental data, their predictive power, however, is limited by the accuracy of the proteins' connectivity in the signaling pathways used as a scaffold, on top of which the models are built. Thus, if the signaling pathways (as obtained from the literature) represent the proteins' connectivity inaccurately, then the resultant models will yield erroneous predictions. To capture the true signaling motifs of the interrogated cell type, experimental data are usually incorporated and a training algorithm is employed to calibrate the model to best fit the data at hand. On this front, significant work has been published using (i) optimization algorithms, such as genetic algorithms or regular optimization formulations and (ii) sensitivity analysis.

Regarding optimization formulations, in the work by Saez-Rodriguez *et al*. [28,29], a signaling pathway was put together downstream of six receptors of interest, based on literature citations of protein interactions, and Boolean logic was used to model signal transduction in the pathway. Then, a genetic algorithm pruned the pathway by removing reactions that seemed to contradict high throughput experimental data. In the work by Mitsos *et al*. [30–32], an Integer Linear Programming formulation was introduced to prune the pathway so that it best fits the experimental data at hand, diminishing the required CPU time, thus, allowing the interrogation of complex pathways and phosphoproteomic datasets. An ILP formulation was also used in the work by Melas *et al*. (Detecting and removing inconsistencies

between experimental data and signaling network topologies using integer linear programming on interaction graphs, accepted in *PLoS Comp Biol*, 2013) to identify and remove inconsistencies between signaling pathway topologies and phosphoproteomic data. In addition to pruning the network, two more strategies were employed: (i) addition of *de novo* reactions and (ii) identification of minimum correction sets, defined as the minimum set of nodes that have to be corrected to obtain a perfect fit of the data. Other than optimization algorithms, methodologies based on sensitivity analysis are also used [33]. In these approaches, the connectivity of the proteins in the signaling pathway is inferred by considering infinitesimal changes in the activation of an arbitrary node A and monitoring changes in the activation of node B, while keeping the activation of other nodes constant. If the two nodes are co-regulated then an interaction may be present between them.

Even though a few years ago the phosphoproteomic data needed to perform this type of analysis were not easy to obtain, recent advancements in high throughput proteomics technologies now allow the quantification of dozens of proteins per sample with minimum sample requirements. The three platforms most suitable for this endeavor are protein microarrays (or planar arrays) [34,35], the xMAP technology (or suspension arrays) [36] and flow cytometry [37] since they combine high signal and sample throughput (see Box 1). MassSpec on the other hand, is ideal for exploratory purposes (identify which proteins are expressed in a specific tissue etc.) but cannot be used in the clinic because of its sample requirements.

The systematic modeling of signaling pathways, as implemented in the discovery phase, results in predictive models of the signaling mechanisms in the cell type/tissue of interest. In particular, the various data inference algorithms and logic modeling, via leveraging high throughput proteomic data, succeed in integrating in predictive models a multitude of pathways, responsible for most major cellular functions. If applied within a systems pharmacology framework, these approaches could potentially uncover the signaling processes that take place in the patient and provide a systems framework for the interpretation of PD biomarker data. For example,

instead of only measuring the activation levels of a few biomarkers relative to the interrogated disease, without truly identifying the mechanisms that regulate their expression, one could also measure phosphoproteins that play a central role in cellular functions and construct extensive models that correlate them with the predefined biomarkers. In this way, a broader view of the mechanisms that regulate the expression of these biomarkers is obtained and the processes that govern clinical response may be deconvoluted. For example, Iadevaia et al. [38] constructed an ODE model based on phosphoproteomic data, to model signal transduction downstream of the IGF1 receptor and was then leveraged to identify optimal drug combinations for inhibiting cell proliferation. In addition to PD modeling, these approaches can also be used in personalized medicine. The construction of patient specific models may uncover mechanisms in the progression of the disease that differ from patient to patient, facilitating the selection of optimal therapies. On this front, significant work has been published using reverse phase protein arrays (RPPA) [39,40]. For example, Wulfkuhle et al. [41] studied the role of ERK1/2 pathway in ovarian cancer, demonstrating that patterns in signaling pathway activation in ovarian tumors may be patient-specific rather than stage-specific. Moreover, in the work by Ihle et al. [42], RPPA data were used in combination with transcriptomic data to study the effects of KRAS substitutions to protein behavior and how that affected signaling and clinical outcome. Finally, pathway modeling, applied longitudinally during the course of treatment, may uncover mechanisms of drug resistance [43] and facilitate the selection of the optimal frequency of administration.

Identification of drug mode of action (MOA)

Identification of drug mode of action (MOA) refers to the process of understanding how a drug affects signaling activity. Even though the binding affinities of the drug are well known from bioactivity assays performed at earlier phases of drug development, the functional effects of the drug on the signaling mechanisms of the target tissue are usually not fully characterized. As a result, off target effects may still be identified several years after the drug has been made available to patients. A number of methods have been proposed for the identification of drug MOA that either employ extensive phosphoproteomic measurements to capture directly drug effects on the signaling level, or exploit public repositories, screening the effects of hundreds of drugs on cell lines and then use a machine learning algorithm to deconvolute these signatures and identify where the interrogated drug acts. Unfortunately, these approaches are generally not part of clinical development. In clinical trials the effects of the drug are usually evaluated with respect to a set of predefined PD and safety biomarkers and/or well-established clinical endpoints. However, adverse drug effects of low incidence may stay undetected until enough patients have been exposed to the drug. In the following paragraphs we present commonly used methodologies for the identification of drug MOA that could be leveraged within a systems pharmacology framework with great benefits.

There are two classes of methods that have been proposed for the identification of adverse drug effects: (i) machine learning approaches, (ii) mechanistic approaches [44]. Machine learning approaches leverage extensive data repositories, capturing the effects of hundreds of compounds on different cell lines/patients, and a clustering or classification algorithm is employed to predict the effects of new (uncharacterized) drugs based on their signature and known targets of similar drugs [45–47]. Any type of data can be used in a machine learning framework including phenotypic, transcriptomic, signaling, chemoproteomic, structural, or other data, as long as there are enough drugs available in the training dataset to guarantee the statistical significance of the predictions. For example, Campillos et al. [47] used drug side effects, as these are listed in package leaflets, to identify more than 1000 drug–drug relations between 746 marketed drugs. In this context, drug–drug relations refer to data driven predictions that show when two drugs share a target. Interestingly, 261 of these relations are between chemically dissimilar drugs, while a number of them were validated experimentally. In another application of machine learning approaches, Iorio et al. [46] exploited the cMAP repository [48], screening the effects more than 6000 compounds on the transcriptomic level of cell lines, to identify

more than 40000 relations between 1300 drugs. Iorio *et al*. used a distance metric to score the similarities in drug signatures and when two drugs demonstrated a statistically significant similarity, a drug–drug correlation was introduced. In another work by Gregori-Puigjané *et al*. [45], a chemoinformatics approach was used leveraging the DrugBank.ca repository to identify previously unreported mechanisms of action targets for drugs.

Even though machine learning approaches make robust predictions, they do not provide insight into the exact mechanism of action of the interrogated drug. This is an inherent limitation of the use of gene expression data, phenotypic, structural, or other data in drug development, since most of the drugs act on the phosphoproteomic level; thus, any effects on the gene expression or other level are second order (i.e. indirect) effects [49]. Mechanistic approaches are based on phosphoproteomic data, circumventing this limitation. Mechanistic approaches use *in vitro* data in the presence or absence of the interrogated drug and either identify their differences, estimating in this manner the drug MOA directly [50–52], or construct models of the signaling pathway before and after drug treatment in order to identify drug induced model alterations. For example, Bantscheff *et al*. [50] used mass spectrometry (MS) to reveal mechanisms of action of clinical ABL kinase inhibitors, including unreported targets of imatinib. Mass spectrometry is the method of choice for these questions outside clinical development, since it allows the complete proteomic profiling of the sample (measuring up to 20000 proteins at the same time), essentially measuring the effect of the interrogated drug on the whole proteome of the target tissue. Unfortunately, MS cannot be used effectively in the clinical phase because of its sample requirements; on the other hand, protein microarrays, Luminex xMAP and flow cytometry can be used. These technologies provide much lower signal throughput compared with MS, however, their output data can be used for the construction of mechanistic models; these models may then be leveraged to identify drug effects on the signaling level. For example, Mitsos *et al*. [30] used Luminex xMAP to construct signaling models of HEPG2 cells upon treatment with four different hepatocellular carcinoma

drugs and then identified their differences with the control model (specific to HEPG2 cells), thus obtaining topology alterations caused by the drug. In this manner, previously unreported off-target effects were uncovered. Protein microarrays can also be used in this context. Their very low sample requirements make them an ideal approach for clinical applications and they have already been used extensively in the study of signaling pathways. Moreover, in a very interesting work by Kumar *et al*. [53], gene expression data were leveraged using causal reasoning to identify the MOA of a novel AKT kinase inhibitor (GSK690693). Causal reasoning is a methodology that infers patterns in the phosphoproteomic level that best explain the observed changes in the gene expression level [54].

The rigorous study of drug MOA, as presented above, has repeatedly identified unreported drug targets, even years after the drug has made it to the market. If applied in clinical development it could uncover fully the mechanism of action, possibly identifying off-target effects that would otherwise remain unnoticed and, thus, affect the outcome of subsequent clinical trials. In this manner the identification of drug MOA may decrease the attrition rate of clinical trials by advancing drug candidates with minimum adverse effects or help towards personalized medicine, where optimal combination therapies and drug dosage can be devised for each patient. Drug repurposing could also be another application.

Prediction of clinical efficacy and toxicity

Prediction of clinical efficacy and toxicity refers to the process of predicting the clinical response to a drug, in terms of validated endpoints, based on phenotypic, transcriptomic, signaling, chemoproteomic, structural or other (preclinical) data. The prediction of efficacy/toxicity is a key issue in all phases of drug development. In the discovery phase, systems approaches either leverage extensive *in vitro* datasets via machine learning or clustering analysis, in an attempt to predict the clinical efficacy/ toxicity of uncharacterized drugs based on their signatures and known clinical outcomes of similar drugs; or construct predictive models to describe the signaling, metabolic or transcriptomic processes in the cell type/tissue of interest and

uncover how these are altered in disease or orchestrate the expression of key toxicity mediators. In the latter case either *in vitro* or *in vivo* data can be used, with the *in vivo* data demonstrating clear advantages in terms of predictive power. Typically though, *in vivo* data availability is considerably lower than *in vitro* data due to cost.

Regarding the data driven methodologies, in similar fashion to the identification of drug MOA (discussed above) where such methodologies are applied, any types of data can be used in a machine learning or clustering framework to predict the clinical efficacy and toxicity of the interrogated drug based on clinical outcomes of similar drugs. In the work by Barretina *et al*. [55] it was demonstrated how the Cancer Cell Line Encyclopedia (CCLE) can be leveraged to predict drug sensitivities of cell lines according to their genotype, facilitating application to personalized medicine. The CCLE includes the full transcriptomic screening of around 1000 cancer cell lines (fully characterized) and for 500 of these also includes the pharmacological profiling of 24 compounds. It is, thus, a valuable resource for personalized medicine. Another valuable resource in gene expression-driven prediction of efficacy is the Genomics of Drug Sensitivity in Cancer (GDSC) project [56], where the effects of 140 drugs were screened against 1200 cancer cell lines, also including dose response data. Moreover, Iorio *et al*. [46] presented a clustering-based approach to identify suitable drug repositionings, exploiting the cMAP repository of gene expression profiles; while in the work by Wessels *et al*. [57], a clinical pharmacogenetic model was developed based on custom data to predict the efficacy of methotrexate in rheumatoid arthritis, further demonstrating how genomic data can be leveraged to obtain robust predictions of drug efficacy. Apart from genomic data, Xiang-Qun Xie [58] demonstrated how chemoproteomic and structural data in pubChem can be used for virtual screening purposes, including the prediction of efficacy and toxicity.

Regarding mechanistic driven methodologies, these include the construction of predictive models (mostly network models) to either describe the signaling, metabolic or transcriptomic processes in the cell type/tissue of interest and how these are altered in disease; or correlate the drug target with key efficacy and toxicity mediators. On this front, the work by Klipp *et al*. [59], Folger *et al*. [60] and Kim *et al*. [61] demonstrates how signaling and metabolic network models can be used for effective drug targeting. In another interesting work by Hwang *et al*. [62], the authors presented a systems approach to tackling prion disease, also suggesting possible therapeutic approaches. Regarding toxicity studies, Cosgrove *et al*. [63], leveraged cytokine release measurements to associate toxicity in human hepatocytes with signaling network dysregulation.

This type of analysis if applied in the clinical phase in conjunction with standard PD modeling, could provide investigators with valuable predictions on the clinical efficacy and toxicity of the interrogated drug, as well as feedback to research for the identification of new targets.

Conclusion

Computational modeling offers an unmatched solution for the interpretation of extensive datasets, increasingly becoming the norm in drug development, as proteomic technologies generate high content data with minimum sample requirements. Unfortunately, the community still seems skeptical to apply elaborate modeling methodologies, such as those applied in the discovery phase of drug development into the clinic. Here we presented a number of systems biology methodologies addressing: (i) the modeling of signaling pathways, (ii) the identification of drug MOA and (iii) the prediction of clinical drug efficacy and toxicity which are currently applied in the discovery phase that could be leveraged within a systems pharmacology framework with great benefits.
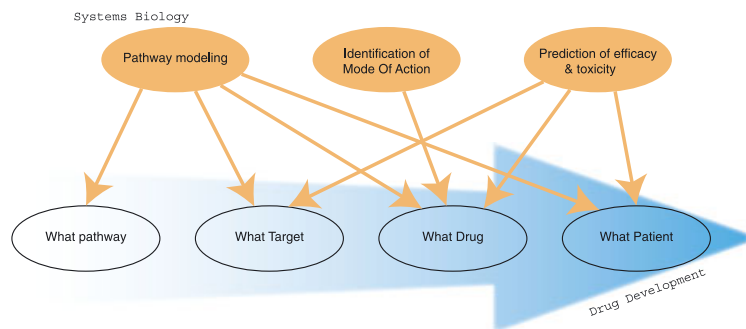
Acknowledgements

Figure 1. Application of the methodologies discussed herein within a systems pharmacology framework to address: (i) the identification of signaling pathways implicated in the interrogated disease, (ii) the identification of optimal drug targets, (iii) selection of optimal drug candidates, and (iv) identification of patient groups for best drug performance

Conflict of Interest

All authors declare no conflict of interest.

References

1. Lalonde RL, Kowalski KG, Hutmacher MM, *et al*. Model-based drug development. *Clin Pharmacol Ther* 2007; **82**(1): 21–32.
2. Tozer TM, Rowland M. Introduction to Pharmacokinetics and Pharmacodynamics: The Quantitative Basis of Drug Therapy. Lippincott Williams & Wilkins, Baltimore, MD 21201 2006.
3. van der Graaf P, Benson N. Systems pharmacology: bridging systems biology and pharmacokinetics-pharmacodynamics (PKPD) in drug discovery and development. *Pharm Res* 2011; **28**(7): 1460–1464.
4. Vicini P, Van Der Graaf PH. Systems pharmacology for drug discovery and development: paradigm shift or flash in the pan? *Clin Pharmacol Ther* 2013; **93**(5): 379–381.
5. Yang R, Niepel M, Mitchison TK, Sorger PK. Dissecting variability in responses to cancer chemotherapy through systems pharmacology. *Clin Pharmacol Ther* 2010; **88**(1): 34–38.
6. Ette EI, Williams PJ. Pharmacometrics: *The Science of Quantitative Pharmacology.* Wiley-Interscience, New York 2007.
7. Danhof M, de Lange ECM, Della Pasqua OE, Ploeger BA, Voskuyl RA. Mechanism-based pharmacokinetic-pharmacodynamic (PK-PD) modeling in translational drug research. *Trends Pharmacol Sci* 2008; **29**(4): 186–191.

8. Rowland M, Peck C, Tucker G. Physiologically-based pharmacokinetics in drug development and regulatory science. *Annu Rev Pharmacol Toxicol* 2011; **51**(1): 45–73.
9. Danhof M, de Jongh J, De Lange ECM, Della Pasqua O, Ploeger BA, Voskuyl RA. Mechanism-based pharmacokinetic-pharmacodynamic modeling: biophase distribution, receptor theory, and dynamical systems analysis. *Annu Rev Pharmacol Toxicol* 2007; **47**(1): 357–400.
10. Aldridge BB, Burke JM, Lauffenburger DA, Sorger PK. Physicochemical modelling of cell signalling pathways. *Nat Cell Biol* 2006; **8**(11): 1195–1203.
11. Ramakrishnan R, DuBois D, Almon R, Pyszczynski N, Jusko W. Fifth-generation model for corticosteroid pharmacodynamics: application to steady-state receptor down-regulation and enzyme induction patterns during seven-day continuous infusion of methylprednisolone in rats. *J Pharmacokin Pharmacodyn* 2002; **29**(1): 1–24.
12. Kholodenko B, Yaffe MB, Kolch W. Computational approaches for analyzing information flow in biological networks. *Sci Signal* 2012; **5**(220): re1-.
13. Ringner M. What is principal component analysis? *Nat Biotech* 2008; **26**(3): 303–304.
14. Janes KA, Kelly JR, Gaudet S, Albeck JG, Sorger PK, Lauffenburger DA. Cue-signal-response analysis of TNF-induced apoptosis by partial least squares regression of dynamic multivariate data. *J Comput Biol* 2004; **11**(4): 544–561.
15. Bansal M, Belcastro V, Ambesi-Impiombato A, di Bernardo D. How to infer gene networks from expression profiles. *Mol Syst Biol* 2007; **3**: 78.
16. Margolin A, Nemenman I, Basso K, *et al*. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 2006; **7** (Suppl 1): 1–15.
17. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP. Causal protein-signaling networks derived from multiparameter single-cell data. *Science* 2005; **308**(5721): 523–529.

18. Hartemink AJ. Reverse engineering gene regulatory networks. *Nat Biotech* 2005; **23**(5): 554–555.

19. Mahdavi A, Davey RE, Bhola P, Yin T, Zandstra PW. Sensitivity analysis of intracellular signaling pathway kinetics predicts targets for stem cell fate control. *PLoS Comput Biol* 2007; **3**(7): e130.

20. Hoops S, Sahle S, Gauges R, *et al*. COPASI, – a COmplex PAthway SImulator. *Bioinformatics* 2006; **22**(24): 3067–3074.

21. Klamt S, Saez-Rodriguez J, Gilles E. Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Systems Biol* 2007; **1**(1): 2.

22. Kauffman SA. Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol* 1969; **22**(3): 437–467.

23. Thakar J, Pilione M, Kirimanjeswara G, Harvill ET, Albert RK. Modeling systems-level regulation of host immune responses. *PLoS Comput Biol* 2007; **3**(6): e109.

24. Helikar T, Konvalina J, Heidel J, Rogers JA. Emergent decision-making in biological signal transduction networks. *Proc Natl Acad Sci* 2008; **105**(6): 1913–1918.

25. Calzone L, Tournier L, Fourquet S, *et al*. Mathematical modelling of cell-fate decision in response to death receptor engagement. *PLoS Comput Biol* 2010; **6**(3): e1000702.

26. Aldridge BB, Saez-Rodriguez J, Muhlich JL, Sorger PK, Lauffenburger DA. Fuzzy logic analysis of kinase pathway crosstalk in TNF/EGF/insulin-induced signaling. *Plos Comput Biol* 2009; **5**(4): e1000340.

27. Morris MK, Saez-Rodriguez J, Clarke DC, Sorger PK, Lauffenburger DA. Training signaling pathway maps to biochemical data with constrained fuzzy logic: quantitative analysis of liver cell responses to inflammatory stimuli. *PLoS Comput Biol* 2011; **7**(3): e1001099.

28. Saez-Rodriguez J, Alexopoulos LG, Epperlein J, *et al*. Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol Syst Biol* 2009; **5**: 331.

29. Terfve C, Cokelaer T, Henriques D, *et al*. CellNOptR: a flexible toolkit to train protein signaling networks to data using multiple logic formalisms. *BMC Systems Biol* 2012; **6**(1): 133.

30. Mitsos A, Melas IN, Siminelakis P, Chairakaki AD, Saez-Rodriguez J, Alexopoulos LG. Identifying drug effects via pathway alterations using an integer linear programming optimization formulation on phosphoproteomic data. *PLoS Comput Biol* 2009; **5**(12): e1000591.

31. Melas I, Mitsos A, Messinis D, Weiss T, Alexopoulos L. Combined logical and data-driven models for linking signalling pathways to cellular response. *BMC Systems Biol* 2011; **5**(1): 107.

32. Melas IN, Mitsos A, Messinis DE, Weiss TS, Rodriguez J-S, Alexopoulos LG. Construction of large signaling pathways using an adaptive

33. Kholodenko BN, Kiyatkin A, Bruggeman FJ, Sontag E, Westerhoff HV, Hoek JB. Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. *Proc Natl Acad Sci* 2002; **99**(20): 12841–12846.

34. Wolf-Yadlin A, Sevecka M, MacBeath G. Dissecting protein function and signaling using protein microarrays. *Curr Opin Chem Biol* 2009; **13**(4): 398–405.

35. Berrade L, Garcia AE, Camarero JA. Protein microarrays: novel developments and applications. *Pharm Res* 2011; **28**(7): 1480–1499.

36. Alexopoulos LG, Saez-Rodriguez J, Cosgrove BD, Lauffenburger DA, Sorger PK. Networks inferred from biochemical data reveal profound differences in Toll-like receptor and inflammatory signaling between normal and transformed hepatocytes. *Mol Cell Proteom* 2010; **9**(9): 1849–1865.

37. Perez OD, Nolan GP. Simultaneous measurement of multiple active kinase states using polychromatic flow cytometry. *Nat Biotech* 2002; **20**(2): 155–162.

38. Iadevaia S, Lu Y, Morales FC, Mills GB, Ram PT. Identification of optimal drug combinations targeting cellular networks: integrating phosphoproteomics and computational network analysis. *Cancer Res* 2010; **70**(17): 6704–6714.

39. Malinowsky K, Wolff C, Ergin B, Berg D, Becker KF. Deciphering signaling pathways in clinical tissues for personalized medicine using protein microarrays. *J Cell Physiol* 2010; **225**(2): 364–370.

40. Yu X, Schneiderhan-Marra N, Joos TO. Protein microarrays for personalized medicine. *Clin Chem* 2010; **56**(3): 376–387.

41. Wulfkuhle JD, Aquino JA, Calvert VS, *et al*. Signal pathway profiling of ovarian cancer from human tissue specimens using reverse-phase protein microarrays. *Proteomics* 2003; **3**(11): 2085–2090.

42. Ihle NT, Byers LA, Kim ES, *et al*. Effect of KRAS Oncogene substitutions on protein behavior: implications for signaling and clinical outcome. *J Natl Cancer Inst* 2012; **104**(3): 228–239.

43. Garraway LA, Jänne PA. Circumventing cancer drug resistance in the era of personalized medicine. *Cancer Discov* 2012; **2**(3): 214–226.

44. Xie L, Xie L, Kinnings SL, Bourne PE. Novel computational approaches to polypharmacology as a means to define responses to individual drugs. *Annu Rev Pharmacol Toxicol* 2012; **52**(1): 361–379.

45. Gregori-Puigjané E, Setola V, Hert J, *et al*. Identifying mechanism-of-action targets for drugs and probes. *Proc Natl Acad Sci* 2012; **109**(28): 11178–11183.

46. Iorio F, Bosotti R, Scacheri E, *et al*. Discovery of drug mode of action and drug repositioning from

transcriptional responses. *Proc Natl Acad Sci* 2010; **107**(33): 14621–14626.

47. Campillos M, Kuhn M, Gavin A-C, Jensen LJ, Bork P. Drug target identification using side-effect similarity. *Science* 2008; **321**(5886): 263–266.

48. Lamb J. The connectivity map: a new tool for biomedical research. *Nat Rev Cancer* 2007; **7**(1): 54–60.

49. Fliri AF, Loging WT, Volkmann RA. Drug effects viewed from a signal transduction network perspective. *J Med Chem* 2009; **52**(24): 8038–8046.

50. Bantscheff M, Eberhard D, Abraham Y, *et al*. Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat Biotechnol* 2007; **25**(9): 1035–1044.

51. Bantscheff M, Hopf C, Savitski MM, *et al*. Chemoproteomics profiling of HDAC inhibitors reveals selective targeting of HDAC complexes. *Nat Biotech* 2011; **29**(3): 255–265.

52. Li J, Rix U, Fang B, *et al*. A chemical and phosphoproteomic characterization of dasatinib action in lung cancer. *Nat Chem Biol* 2010; **6**(4): 291–299.

53. Kumar R, Blakemore S, Ellis C, *et al*. Causal reasoning identifies mechanisms of sensitivity for a novel AKT kinase inhibitor, GSK690693. *BMC Genomics* 2010; **11**(1): 419.

54. Kleinberg S, Hripcsak G. A review of causal inference for biomedical informatics. *J Biomed Informat* 2011; **44**(6): 1102–1112.

55. Barretina J, Caponigro G, Stransky N, *et al*. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 2012; **483**(7391): 603–307.

56. Yang W, Soares J, Greninger P, *et al*. Genomics of drug sensitivity in cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res* 2013; **41**(D1): D955–D961.

57. Wessels JAM, van der Kooij SM, le Cessie S, *et al*. Pharmacogenetics Collaborative Research G. A clinical pharmacogenetic model to predict the efficacy of methotrexate monotherapy in recent-onset rheumatoid arthritis. *Arthr Rheum* 2007; **56**(6): 1765–1775.

58. Xie X-Q. Exploiting PubChem for virtual screening. *Expert Opin Drug Discov* 2010; **5**(12): 1205–1220.

59. Klipp E, Wade RC, Kummer U. Biochemical network-based drug-target prediction. *Curr Opin Biotechnol* 2010; **21**(4): 511–516.

60. Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T. Predicting selective drug targets in cancer through metabolic networks. *Mol Syst Biol* 2011; **7**: 501.

61. Kim HU, Sohn SB, Lee SY. Metabolic network modeling and simulation for drug targeting and discovery. *Biotechnol J* 2012; **7**(3): 330–342.

62. Hwang D, Lee IY, Yoo H, *et al*. A systems approach to prion disease. *Mol Syst Biol* 2009; **5**: 252.

63. Cosgrove BD, Alexopoulos LG, Hang T-C, *et al*. Cytokine-associated drug toxicity in human hepatocytes is associated with signaling network dysregulation. *Mol BioSystems* 2010; **6**(7): 1195–1206.

64. Alexopoulos LG, Saez-Rodriguez J, Espelin C. High throughput protein-based technologies and computational models for drug development, efficacy and toxicity. In Drug Efficacy, Safety, and Biologics Discovery: Emerging Technologies and Tools, Xu JJ, Ekins S (eds). Wiley & Sons, Hoboken, New Jersey 2009.