BLENDING BIOLOGY AND CHEMISTRY TO ENABLE SYSTEMS PHARMACOLOGY

MIQUEL DURAN-FRIGOLA

PhD Thesis in Biomedicine

Supervised by (Director de la Tesi) Dr. Patrick Aloy
Tutored by (Tutor de la Tesi) Dr. Baldo Oliva

Structural Bioinformatics and Network Biology laboratory
Institute for Research in Biomedicine (IRB Barcelona)
Joint IRB-BSC-CRG Program in Computational Biology

Universitat Pompeu Fabra

2011–2015

TESI DOCTORAL UPF - 2015

upf. **Universitat
Pompeu Fabra**
Barcelona

It is only now and then
in some very remote and backward agricultural district
that an antiquarian may still discover a square house.

— Edwin A. Abbott, *Flatland*

Als meus avis, tan diferents.

ABSTRACT

The avalanche of data that followed the sequencing of the human genome has revealed an overwhelming biological complexity. No simple molecular explanation exists for most of the diseases and, in consequence, simple therapies have low probability of success. The emerging field of systems pharmacology seeks drugs of broad impact on molecular networks. To achieve so, it is necessary to integrate heterogeneous data, at different levels of complexity, and find correlations between them. This translational exercise is, perhaps, the major concern of current biomedical research.

In this Thesis we undertake part of this challenge through cases that orbit the drug discovery endeavor. Using computational methods in various areas of bioinformatics and chemoinformatics, we link chemical, biomolecular and phenotypic data to provide a more holistic view of pharmacology.

Keywords: systems pharmacology, network biology, bioinformatics, chemoinformatics

RESUM

L'allau de dades que ha seguit la seqüenciació del genoma humà està revelant una increïble complexitat biològica. No existeix una explicació molecular simple per a la majoria de les malalties i, en conseqüència, les teràpies simples tenen baixes probabilitats d'èxit. L'emergent camp de la farmacologia de sistemes busca medicaments d'ampli impacte en les xarxes moleculars. Per a aconseguir-ho, és necessària la integració de dades heterogènies, a diferents nivells de complexitat, i la capacitat de trobar correlacions entre elles. Aquest exercici translacional és, probablement, la major preocupació de la recerca biomèdica d'avui.

En aquesta Tesi assumim part d'aquest repte a través de casos que orbiten el descobriment de fàrmacs. Mitjançant mètodes computacionals en àrees diverses de la bioinformàtica i la quimioinformàtica, connectem dades químiques, biomoleculars i fenotípiques per a facilitar una visió més holística de la farmacologia.

Paraules clau: farmacologia de sistemes, biologia de xarxes, bioinformàtica, quimioinformàtica

AGRAÏMENTS

Faig aquest apartat curt i fred, tot i sentir-lo llarg i càlid, perquè em limita el llenguatge. Agrair com vull agrair només podré-ho fer personalment, un a un, em temo. He de començar pel meu Director de Tesi, en Patrick Aloy, de qui sempre he rebut suport i reptes, i de qui he après aproximadament tota la ciència que sé. A ell li dec la manera de fer les preguntes i la manera de respondre-les, i l'oportunitat de treballar en un grup ambiciós i competent. He col·laborat amb en Roberto Mosca, que té una precisió extraordinària, l'Andreas Zanzoni, la Montse Soler-López, la Samira Jaeger, amiga per sempre, la Teresa Juan-Blanco, la Lydia Siragusa, en Rodrigo Arroyo i en Víctor Alcalde, uns altres amics eterns. Aquesta Tesi és molt seva. Els altres membres del grup, actuals i passats, m'han donat incomptables idees i diversió, els llisto desordenadament, però a consciència: Manuel Alonso-Tarajano, Laura Isús, Lídia Mateo, Roger Olivella, Amélie Stein, Nahuai Badiola, Roland A. Pache, Eva Capdevila, Isabelle Brun-Heath, Jofre Tenorio-Laranga, José Luis Rodríguez, Arnaud Céol, Eduard Pauls, Sergi Bayod i Jacopo Negroni.

Estem, a més, en un institut de recerca únic. Hi he trobat intel·ligències peculiars: Pedro Sfriso, amb qui he treballat amb estranya sinergia, Carlos Maíllo i Gemma Turon, us auguro el millor. També he fet bons amics: Simone Pieretti, Francisco Freixo, Jessica Vitos, Milos Tatarski, Giulio Chiesa, Mariano Maffei, Pablo Barrecheguren, Anna Montaner, Àlex de la Fuente, Pep Garcia, Salva Guardiola, Benjamí Oller, Federica Battistini, i un llarg i amnèsic etcètera: gràcies per ser-hi, que bé haver-vos conegut. La meva estada a Tel Aviv va ser també un regal, thanks to Eytan Ruppin for the orthogonal thinking and enthusiasm, and thanks to Erez Persi and Matthew Oberhardt for the fun in our projects. Keren Yizhak, Edoardo Gaude, Adam Weinstock, Yoav Teboulle and Rotem Katzir, you all made the stay profound and exciting.

Tampoc m'oblido dels que m'han aconsellat i aprovat durant tot aquest temps: Modesto Orozco, Manuel Pastor, Xavier Barril i Baldo Oliva, heu sostingut el meu entusiasme pel projecte. L'administració de l'institut i el finançament del Ministeri, EMBO i FEBS han fet que tot plegat fos més pràctic.

Finalment, i principalment, em dirigeixo a la família i als amics que són família. Pare, mare, germà, avis, tota la resta, Ferran Roure, Mireia Bosch, Pol Duran, Joan Garcias, Sílvia Vilaprinyó i Eudald Díaz: mai llegireu aquesta Tesi, però l'he escrit per a vosaltres.

PREFACE

The field of biomedicine was peaking in my hometown Barcelona by the time I finished the chemistry degree, back in 2009. A number of research institutes had been created in the years preceding the economic crisis, and Barcelona was determined to be the reference spot for biomedical sciences in Southern Europe. Relevant discoveries were day in day in the media and in glamourous scientific journals, and one could feel a unique enthusiasm for biological innovation.

I have to admit that, at the time, I had some reservations and prejudices about biology in general, and molecular biology in particular. My naïve perception was that it was a merely descriptive discipline, and being a chemist initialized in quantics I was most of all interested in fundamental laws —"all science is either physics or stamp collecting", said Lord Rutherford. I suspect, though, that the mere popularity of biology was in a way attractive to me: after all, it took me a couple of years to figure out my scientific itinerary. I remember myself doing a lot of reading and, rather slowly, I appreciated an increasing use of mathematics and computation in biology. It was apparent that the impressive amount of data that was being generated required some shifts in the traditional biological thinking. I understood that this would offer many opportunities, and I found that, being a chemist with some programming notions and an unmet passion for statistics, bioinformatics was the natural entry point to biomedicine.

This Thesis is the result of my first steps into computational biology, with a marked focus on pharmacology to somehow recycle my chemical background. The Reader will notice that this is, above all, an exploratory work, as my experience has been, composed of several questions at the edge of chemistry, biology and medicine. I hope that the text is able to connect the very distinct concepts, and that it does not read erratic. It is my greatest desire that some of the rough ideas that we propose herein are continued some day, hopefully to contribute to the development of better drugs.

CONTENTS

Part I

INTRODUCTION

GENERAL INTRODUCTION

1.1 THE COMPLEXITY OF HUMAN DISEASE

Human diseases emerge from the malfunctioning of complex biological systems [1, 2, 3]. The cellular machinery is interconnected, and the impact of a specific genetic alteration is seldom confined to the gene that carries it, but propagates and alters the activity of gene products that otherwise carry no defects [4]. Even for Mendelian disorders, where genetic abnormalities can be traced in family pedigrees, it is now recognized that focusing on a single gene yields too simplistic pictures [2]. For more complex disorders —including cancer, diabetes, autism, and obesity— individual genetic factors cannot even be pinned down, since these diseases are caused by perturbations of intracellular and intercellular networks [5, 6]. Due to the combined effect of multiple genetic determinants, the variability between individuals and the importance of environmental factors, the phenotype of complex diseases cannot be easily determined from the genotype.

Unveiling the links between genotype and phenotype is at the core of biomedicine, a field that has been revolutionized by the application of high-throughput technologies following the first drafts of the human genome, back in 2001 [7, 8]. The long series of biological events that associate a genetic risk variant with the development of a disorder can be monitored through a range of biochemical intermediates, including different types of RNA, proteins, and metabolites [1]. Today, an overwhelming amount of transcriptomics, proteomics and metabolomics data, among others, are being published, accompanied with analyses that report lists of genes whose mutations correlate with a disease. This mainstay approach to genetic studies, based on the independent discovery of genes, is a critical and valuable step, yet it is fundamentally insufficient to describe the intricate architecture of disease pathogenesis [9].

1.1.1 *The advent of systems biology*

The emerging field of systems biology is taking over the challenge to integrate and digest the available molecular data. This in-vogue field has been proclaimed a paradigm shift by many [10], since once for all it faces the intricate network of interactions that exist between the constituents of the cell. It is probably an overstatement to say that systems-level, integrative reasoning is new to molecular biology [11]. Outside the mainstream reductionist school, nonequilibrium thermodynamics theory flourished in the 1930s to set foundations on molecular coupling [12, 13]. Biochemical processes need to produce entropy as a driving force, and an early noticed paradox was how to explain the or-

"...the phenotype of complex diseases cannot be easily determined from the genotype."

dering that occurs in developmental biology when entropy must be increased. A compelling explanation was that ordering processes are concatenated with chaotic ones, i.e. that biology requires interactions [14] —nonequilibrium thermodynamics constituted a prelude to modern systems biology, but in the 20th century the seminal studies in small systems did not get much attention (Figure 1A).

Instead, reductionism has dominated research in molecular biology, providing an enormous supply of knowledge about individual cell components. Automation and miniaturization of molecular assays triggered the expansion of the field of bioinformatics in the late 1990s, which is also rooted in the reductionist view [11]. Early bioinformatics, although large-scale in nature, was mostly focused on statistical models and lacked a mechanistic framework, i.e. the virtue to integrate knowledge. Systems biology emanates from bioinformatics and takes the whole genome, including its interactions, as the subject of study. To formalize the properties of cellular processes, systems biology incorporates principles from physics, chemistry and biology, which explains why multiple disciplines converge in current biomedical research. To contextualize these cellular processes, broad maps of macromolecular interactions are necessary. One of the most urged tasks in current biology is to unveil these maps and understand the signals that are transmitted throughout [5].

"Systems biology (...) takes the whole genome, including its interactions, as the subject of study."

1.1.2 *The human interactome*

A convenient assumption in systems biology is that the many components of the cell communicate with each other through binary interactions. At an abstract level, complex systems can be seen as a network, or graph, with nodes representing biological entities, and edges, the interactions between them (Figure 1B). A popular implementation of this network view comprises the set of physical protein-protein interactions (PPIs). Of the varied methodologies that can detect PPIs, two are currently in wide use for high-throughput discovery. Binary interactions are primarily identified by ever-improving versions of the yeast two-hybrid (Y2H) system, and identification of protein complex membership is performed by affinity- or immuno-purification to first isolate protein complexes, and then by mass spectrometry (AP/MS) to detect their constituents. In the recent years, there has been an exceptional growth of PPI data, yielding interactome maps for several organisms, including human (Figure 1B) [15].

Network topology

The vast interactome data available in the public domain is leveraged by advances in network analysis [16, 17]. A breakthrough in network theory was the realization that systems' architecture is sustained by a few simple properties that are persistent across most networks of technological and scientific interest. Much like social, computer or semantic networks, PPI networks are characterized by a set of organizing principles that are far from random, confer-

ring them with a series of contextual properties and patterns. For example, an unexpected property of biological networks is the existence of a few highly connected nodes, called hubs, that interact with a majority of peripheral entities. More concisely, in these so-called 'scale-free' networks the degree distribution of node connectivity approximates a power law (Figure 1B) [18]. A noticeable consequence of the scale-free property is the existence of key genes that hold the network together. Accordingly, hubs in PPI networks are a constant focus of attention, and claims are made as to their tendency to correspond to essential [19] or disease-related genes [20]. Indeed, in several model organisms protein connectivity and other network centrality measures appear to be relevant to biology, yet suspicion persists mostly due to research biases and low coverage of interactomes [21]. In the case of the human interactome, the current coverage is estimated to be around the 20%, which means that analyses are conducted on networks that miss 80% of the edges [4].

Disease-related modules and robustness

Despite their incompleteness, cumulative evidence shows that PPI networks are globally able to capture biological features, particularly if the lack of data is taken into account using percolation theory, that treats the current interactome as a sample of the real one [4]. With this rationale, it has been shown that proteins involved in the same or related biological processes tend to interact with each other. This observed locality, or modularity, is a backbone of network biology [2], and suggests that disease-related genes are placed in certain vicinities that are tightly linked to the pathogenic process.

In general, biologically-relevant modules are found in the network by calculating functional enrichments on densely connected neighborhoods, or by locating several disease-related genes and checking their relative distance [2]. Functional and disease modules overlap only to a certain extent, since many complex diseases comprise multiple cellular functions: modularity is thus able to account for complex procedures and relate them. Today, a global map of human diseases (including hundreds of conditions and their molecular proximities) is available thanks to the network-based analysis of otherwise poorly informative gene lists [4].

It is thus apparent that biological processes are well organized by the interactome. A close-up look into this organization has revealed a remarkable degree of biological robustness, with scale-free, redundancy and sparseness as major contributing network properties [22]. Indeed, biological systems sustain relatively normal behavior upon e.g. most genetic mutations, regime shifts in the physical environment, or stochastic fluctuations in molecular concentrations [23, 24, 25]. The unique capacity of biological networks to elicit robustness, which in turn is associated with homeostasis and evolution, suggests that the currently available interactome is sufficient to capture the global architecture of the cell.

Interactomes are mostly descriptive

The current challenge in interactome analysis is to move beyond the mere delineation of cell's architecture. PPI network analysis is mostly topology-based, and eminently statistical, providing no quantitative outcomes. To be truly predictive, cellular networks require that both the magnitude and dynamics of the interactions are considered. This is not feasible at large for PPI networks, because high-throughput experiments only detect the presence or absence of physical interactions, and it is challenging to derive their function and understand signal transmission. Fortunately, once the interactomes are better annotated, systems biology is prepared to produce quantitative tools —recent advances in smaller, well-characterized systems such as metabolism are setting the bases of genome-scale predictive models [26].

1.1.3   *Genome-scale metabolic models*

Metabolism is determined by genetics and environment. Models of human metabolism exist that take these factors into account thanks to the reconstruction of metabolic networks and analysis of reaction flows, providing the most mature and predictive systems biology tools available to date [26].

Metabolic reconstruction

Different to high-throughput PPI interactomes, metabolic networks are built from the available annotation of the human genome, together with the collective knowledge on enzymes, transporters and metabolites that have been studied for decades. Component-by-component reconstruction of genomic and literature data has produced curated metabolic networks for many organisms [27, 28]. In these networks, nodes represent metabolites and edges are reactions annotated with the corresponding metabolic genes. For human, several reconstructions are available, containing thousands of enzymes and metabolites, and some of them being specific to certain tissues [29, 30].

Constraint-based modeling

Reconstructed genome-scale metabolic networks may be converted to a mathematically consistent format, the stoichiometric matrix (Figure 1C). This matrix is the central component of a constraint-based model, which can be queried by a myriad of computational techniques —flux balance analysis (FBA) is perhaps the most prominent one [31]. FBA calculates the flow of metabolites through the network, thereby making it possible to predict metabolic states of physiological relevance. In FBA, the tabulation of stoichiometries imposes constraints on the system, ensuring that the total amount of metabolites produced is equal to the amount being consumed at steady state. Reactions can also be given upper and lower bounds to delimit their allowable fluxes. These stoichiometries and bounds constrain the space of feasible flux distributions of the system, i.e. the rates of consumption or production of every metabolite. More importantly,

some reactions can be linked to phenotypic traits. Biomass production, for instance, correlates with cellular growth [32] and proliferation [33]. In a metabolic model, biomass production is just a compendium reaction that describes the rate at which metabolites are converted into biomass constituents such as nucleic acids, proteins, and lipids. Thus, genome-scale metabolic models provide a genuinely systemic link between molecular traits and phenotype, and demonstrate that quantitative modeling is possible on large biological networks.

1.2 THE SIMPLICITY OF MODERN DRUGS

Overall, the ever-growing pile of systems biology studies is moving our understanding of human disease towards a holistic view, where cellular components interplay to perform biological processes. This recognized complexity does not match the apparent simplicity of current drugs, which have long been optimized under the 'one disease, one target, one drug' rationale [34]. This notion, although alluring, has yielded disappointing results, and its fundamental flaws are a matter of active debate [35].

1.2.1 *The decline of productivity in the pharmaceutical industry*

In the quest for single molecular targets, pharmaceutical research has experienced a downturn that goes beyond demand and competition [36]. Arguably, this decline is also due to an increasing concentration of R&D investments in areas with high risk of failure, which correspond to pending therapeutic needs and uncharted biological mechanisms. There is the perception that no easy targets remain, and incremental discovery in exploited therapeutic areas is discouraged [37]. While the cost of developing a new drug has dramatically increased, new molecular entities are released at best at constant pace, and attrition rates have risen markedly, especially in late clinical phases [36].

To pursue therapeutic innovation, drug discovery has typically focused on proteins with a known key role in disease pathogenesis. Advances in molecular biology are constantly flagging targets, yet most of them await validation [38, 39, 40]. The target-centered approach is very attractive since it sets a rational basis for the development of new medicines, with a specific biological hypothesis and a starting point for the identification of ligands. Enormous technical achievements have been made in the identification of targets and compounds that interact with them. This notwithstanding, it is rather perplexing that, in a time dominated by target-centered research, phenotypic screening is still yielding most of the first-in-class medicines [37]. Unfortunately, phenotypic screens are usually insufficient since they are opaque to the molecular mechanism of action, have lower throughput, and involve a blind, erratic optimization of lead compounds.

"…drug discovery has typically focused on proteins with a known key role in disease pathogenesis."

On the contrary, a major shortcoming of the target-centered approach is that the known molecular mechanism may not achieve satisfactory therapeutic in-
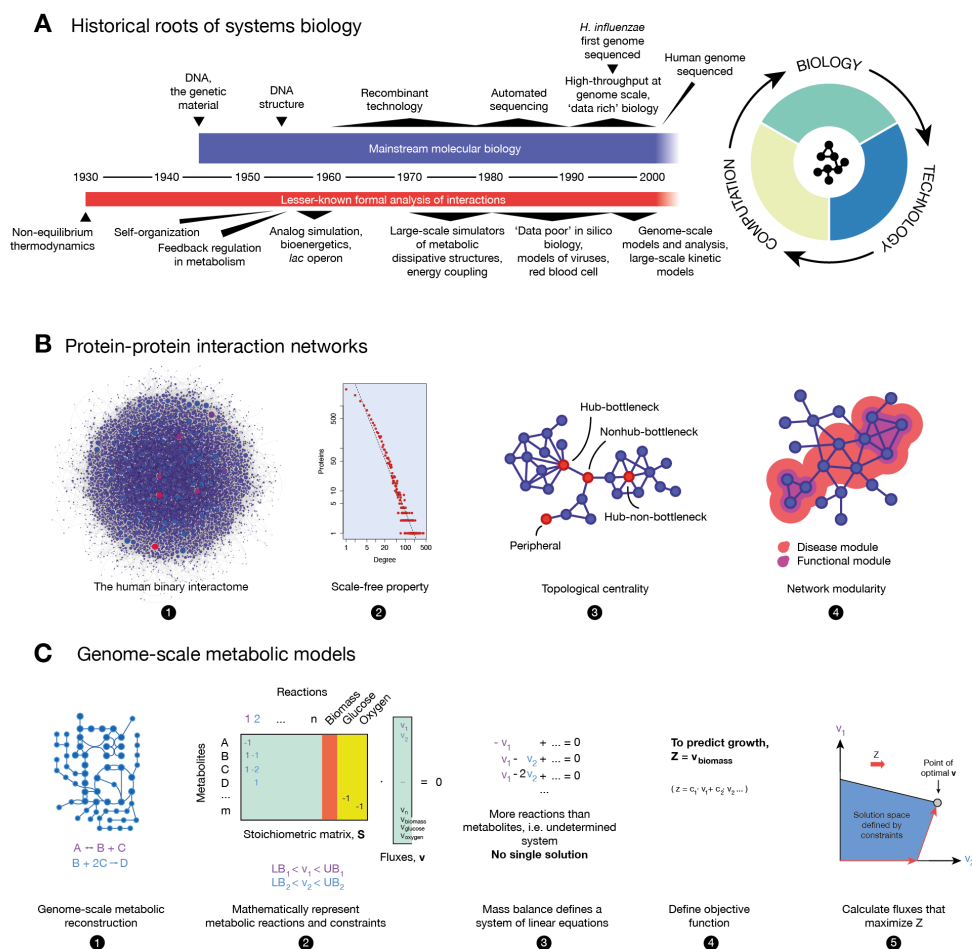
Figure 1: Overview of systems biology. (A) The mainstream origin of systems biology (up) and a lesser-known effort focused on the formal analysis of functional states that arise upon molecular interactions (bottom). Scientific breakthroughs are landmarked in both timelines, leading to the emergence of systems biology in the last decade (adapted from [11]). Systems biology integrates biological knowledge, big data and computation. (B) Protein-protein interaction networks. The (1) binary human interactome (2014) is shown, and its corresponding (2) degree distribution approximates a power law. We also show sketches of (3) node centralities (adapted from [2]), and (4) network modules (adapted from [1]) where the disease module spans two functional modules. (C) Flux balance analysis in genome-scale metabolic models. (1) The metabolic network can be expressed as a set of stoichiometric reactions that (2) can be represented in a matrix format (S), together with a flux vector (v) to which lower and upper bounds (LB and UB) can be imposed. The steady-state assumption dictates that $S \cdot v = 0$, which yields (3) an undetermined system of equations. (4) Biomass production can be chosen as an objective function Z, and (5) maximized to predict the physiological fluxes of each of the reactions.

dices due to the underlying complexity of disease pathogenesis [41]. Another equally important factor, at the very detailed level, is the poor understanding

of the specific molecular interactions that occur between a drug and its target. There are many facets of these interactions that determine the therapeutic outcome: sites of binding, affinity and kinetics, functional impact and specificity, all add to the ultimate pharmacological response. It has been argued that a focus on target-based drug discovery, without sufficient recognition of the fine molecular details, could largely contribute to the high attrition rates [37]. This is often an underappreciated challenge, being drugs almost uniquely optimized for binding affinity [42].

1.2.2 *Protein structures to facilitate drug design*

Structural biology can fulfill the deficit of information between identified hits and the many criteria that must be met to convert them into preclinical candidates [43]. The utility of 3D structures for drug discovery was appreciated short after the first globins were crystallized [44, 45] —even if X-ray structures of drug targets were usually not available, hints could be obtained through comparative modeling of homologs [46]. Additionally, topographies of ligand-target complementarity have been exploited to optimize potency and selectivity ever since. Despite these applications, for years structural biology has been a valuable yet not critical complement to pharmaceutical research [43].

Structural biology at large scale

Finally, the omics revolution has brought structural biology at the forefront of drug discovery, beyond its classical role in lead optimization [43]. Structural genomics aims at automating all steps in protein crystallography to keep structure determination in pace with the large amount of sequences that are being released [47, 48, 49]. High-throughput crystallography is unveiling representative structures for many families in several genomes, and homology modeling is reaching ideal accuracies [50, 51]. As more structures are resolved, it becomes more important to promptly annotate them. In particular, the identification of ligand binding regions that could be involved in productive intermolecular interactions is paramount, and guides both virtual screening and *de novo* drug design, where small molecules are placed in the binding cleft of protein targets, and spatial, electrostatic, van der Waals, and hydrogen bonding interactions are evaluated [52, 53]. Related to this precise analysis, the design of specialized chemical libraries is now feasible due to atomistic details, moving away from the diverse but sparse libraries of synthesis produced by standard combinatorial chemistry [54]. Large-scale structural biology is thus permeating to the very first steps of the drug discovery pipeline, being the initial selection of a target heavily influenced by structural information as an indicator of the suitability of a protein family to inhibition by small molecules.

"As more structures are resolved, it becomes more important to promptly annotate them."

Difficult targets

To guide target selection, retrospective analysis of successful drug campaigns pointed out features that make some proteins 'unligandable', e.g. featureless

binding sites, scarce hydrogen-bond donors and acceptors, metal ions, adaptive changes in conformation, and lipophilicity at the protein-ligand interface [55]. A majority of the putative targets that are underlined in disease genetics studies do not seem to be tractable [56]. Of notice, protein-protein interfaces have several troublesome features, and attractive structural regions such as allosteric sites are many times transient and only exposed in certain protein conformations [55]. Nowadays, there is intense research to mitigate the belief that most targets are unligandable, and game-changing technologies like fragment screening are expanding our chemical toolbox [57].

1.2.3   *The chemo-centric view of pharmacology*
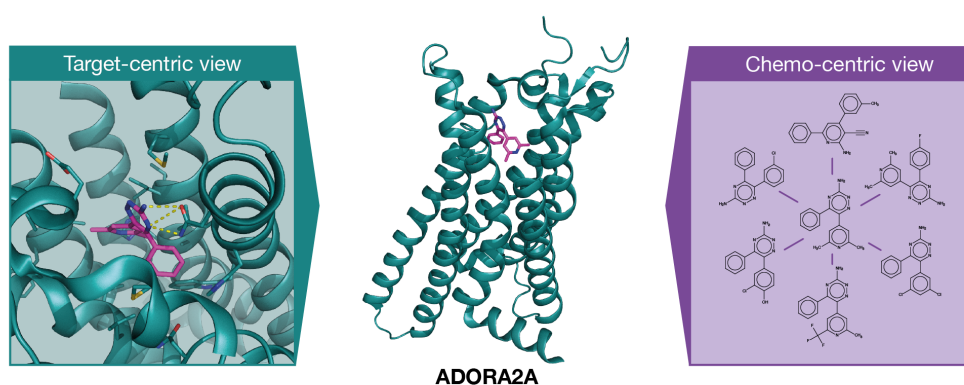
Figure 2: The protein and the small molecule structural viewpoints. Target-centric and chemo-centric views of the Adenosine A2A receptor (ADORA2A) modulation by 6-(2,6-dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine (T4G). On the left, the key molecular interactions in the T4G-ADORA2A binding. On the right, T4G derivatives that also bind to ADORA2A below 10 μM.

With such an ubiquitous role of structural biology in modern drug discovery, it is difficult to conceive an endeavour exempt of thorough knowledge on the intended drug targets. However, before the molecular biology wave, pharmacologists were almost unaware of any molecular detail, and worked with a radically different approach that has been progressively abandoned [58, 59]. Rather than target-centric, classical pharmacology was chemo-centric (Figure 2). Even when the molecular targets where roughly known, like in the case of G-protein coupled receptors, the main goal was to classify them based on the chemistry of the active molecules. These efforts led to rich chemical taxonomies that correlated well with bioactivity. Most times, molecules were directly tested in phenotypically-relevant systems such as organs or animal models, and this trial-and-error strategy yielded remarkable medicinal discoveries. Unfortunately, the chemo-centric picture carried major limitations, most notably a lack of a theoretical background and the need for previous bioactivity data, which made it unable to catch up with the large collection of new targets that were released after the genomics revolution.

"...before the molecular biology wave, pharmacologists were almost unaware of any molecular detail..."

The acknowledgement that the one-disease-one-gene-one-drug paradigm is also limited in many aspects encouraged the drug discovery community to revisit some old ideas and implement them with modern technologies. Nowadays, cell-based phenotypic screenings are performed with high success [37], overlooking the individual targets, and the chemo-centric classification of proteins is broadly feasible thanks to the explosion of chemogenomics databases [59, 60, 61]. Enough bioactivity information is available to characterize thousands of proteins in the light of their ligands. From this angle, with origin in classical pharmacology, two proteins are similar if they have similar ligands, providing a complementary view to evolutionary and structural classifications. This notion has been applied with notable results to target and off-target detection [62], even if it is largely dependent on the amount and quality of the assay data: the approach is sensitive to biases in chemical libraries [63].

Nevertheless, the current availability of chemical library bioactivities is boosting small molecules in a time of big data dominated by bioinformatics. Yet, public chemoinformatics tools and standards lag behind bioinformatics; the main reason for this is that bioinformatics and genomics initiatives have their roots in the academic environment, while most chemical screens are conducted in the private sector. The crisis in the pharmaceutical industry, together with the widespread belief that small molecules should be reincorporated into biological models, is encouraging a unification of efforts. Nowadays, chemical and biological data are merged into databases, and sometimes linked to phenotypes [64, 65]. Similar to the advances in the study of disease genetics, the immediate challenge in drug discovery is to push the long lists of independent bioactivity assays and correlations into systems biology frameworks.

1.3 BRIDGING LEVELS OF COMPLEXITY

The problematic tackling of complex diseases with selective therapies suggests that drug discovery should move beyond the limits of reductionism. As nicely put by H. L. Mencken: "For every complex problem there is an answer that is clear, simple, and wrong." Before disease complexity was broadly perceived, in the 1990s, the Nobel Laureate Paul Ehrlich imagined an ideal therapy, a 'magic bullet' exquisitely directed to an invader [66, 67]. Penicillin, perhaps the first great drug, was a magic bullet because it was exceptionally safe and effective for killing bacteria. The switch from mechanisms to targets in drug discovery has adapted Ehrlich's ideal in a strict molecular context: a magic bullet should interact with one selected target, trigger a premeditated response, disengage from the target, and be metabolized and excreted without further effects on the human body (Figure 3). This reasoning has led to rational drug development strategies, enabling the incorporation of physical and chemical cognizance into disease biology, which is crucial to aid molecular design.

In time, omics sciences are revealing astonishing biological complexity, and it is becoming clear that entire systems cannot be reduced to the qualities of the
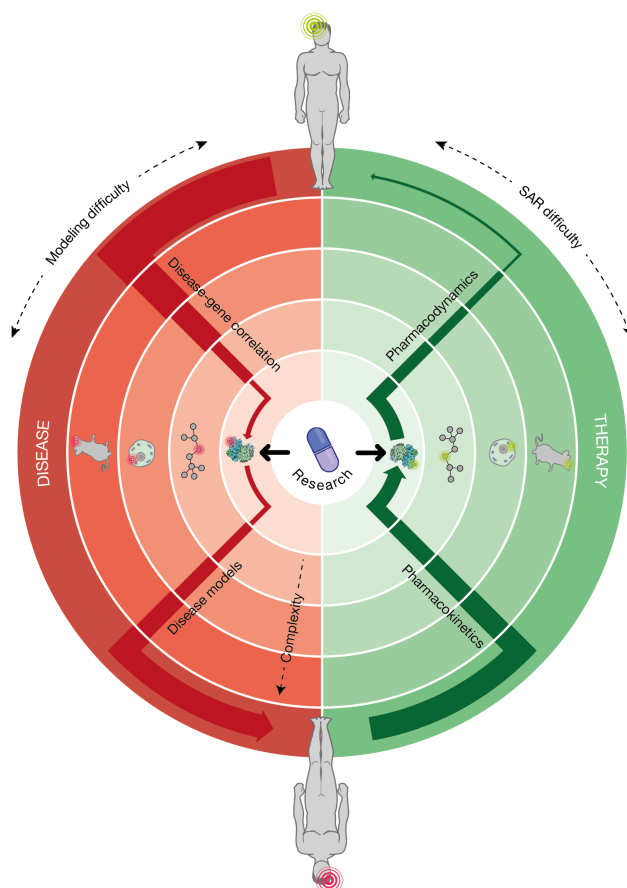
Figure 3: The reductionist approach to disease and therapy. The red half of the circle corresponds to the disease process, and the green half to the therapy. The arrow describes the scientific strategy that is undertaken in order to first understand the disease (red) and then revert it (green). The thickness of the arrow is proportional to the amount of complexity that is accounted for, being inner circles more reduced models, from single proteins to animal models through biochemical pathways and cell assays. In a classical disease genetics study (red), one would aim at finding a causative gene (disease-gene correlation) and, once the gene is identified, it will be contextualized in several disease models of increasing complexity in order to confirm the relevance of the gene and extract mechanistic knowledge. Potentially, this gene could become a candidate drug target (green). Drug discovery is focused on finding molecules that will bind this target, i.e. research is conducted at the inner circle, where structure-activity relationships (SAR) are most simple and easier to optimize. In the physiological context, for the drug to reach the target, pharmacokinetics will have to be controlled, but even if the modulation of the target is accomplished, the complexity of the organism will determine the therapeutic efficacy (pharmacodynamics).

individual parts. Biology has a hierarchical organization, with emergent properties, and downward and upward causation effects. The behavior of a cell, for instance, is controlled both by the properties of its biomolecules and by the properties of its corresponding organ [9]: it appears that a complete picture can

only be obtained by studying the system as a whole and each of its constituents in detail. In reductionist thinking, the natural way to account for this is by isolating biological entities with progressive levels of complexity [68]. Targets, cell cultures, or mouse models are all attempts to simplify the human body (Figure 3). However, putting the object of study in an artificial experimental setup is not neutral and might lead to skewed results. More critically, travelling between orders of complexity is fundamentally impossible from a reductionist, deterministic point of view [9]. In particular, predicting the propagation of a perturbation from one level to another one is a challenging task, since complex systems tend to adapt to environmental conditions while, sometimes, they respond dramatically to small fluctuations.

1.4 THE THESIS INTO CONTEXT

Modern pharmacology has attempted to translate observations from one level of complexity to the other with moderate success. Along the process, massive knowledge on small molecules, genes, and diseases, has been gathered in experiments that range from single target binding assays to post-marketing pharmacovigilance. By nature, these data are diverse and involve fields that are disparate in the scientific mindset. It is the main goal of this Thesis to open new methodological avenues and establish links between multiple data types. Given the heterogeneity of the sources, we use a varied repertoire of computational methods. Our work here has no tool, disease, drug, or gene of interest. Instead, in a rather exploratory manner, we undertake to provide generic solutions and answers to several case examples, with a focus on drug discovery and disease complexity.

"Our work here has no tool, disease, drug, or gene of interest."

Accordingly, the body of the Thesis is a compendium of articles, each of them addressing a different data integration question. In Figure 4, we contextualize the articles in the disease-therapy scheme presented in Figure 3. **To this very general introduction, the Reader can add the individual introduction of each of the articles**.

"…the body of the Thesis is a compendium of articles, each of them addressing a different data integration question."

In brief, the Thesis is rooted in Article 1 (Article 8 and Article 9), where we use drug side effect information to gain mechanistic insights and identify liable chemotypes. The chemo-centric branch of this work is continued in Article 2, and the phenotype-based aspect brought to the cellular level in Article 3 and Article 4, where cancer cell lines are explored. The latter work, in particular, introduces systems biology, and the links between complex networks (Article 10 and Article 11) and protein structures are reviewed in Article 5. Here, we highlight the importance of fine molecular details to enable systems-relevant drug design. Towards this direction, Article 6 presents a method to systematically find functional protein conformations, and Article 7 attempts multi-target perturbations in biological networks.

Articles in the Main Text

1 Analysis of chemical and biological features yields mechanistic insights into drug side effects

2 A chemo-centric view of human health and disease

3 Drug sensitivity in cancer cell lines is not tissue-specific

4 Harnessing cancer's pH-vulnerability to identify an optimal two-step therapeutic strategy

5 Structural systems pharmacology: the role of 3D structures in next-generation drug development

6 Residues coevolution guides the systematic identification of alternative functional conformations in proteins

7 Detecting similar binding pockets to enable systems pharmacology

Articles in the Appendix

8 Recycling side effects into clinical markers for drug repositioning

9 IntSide: a web server for the chemical and biological examination of drug side effects

10 Charting the molecular links between driver and susceptibility genes in colorectal cancer

11 Systematic identification of molecular links between core and candidate genes in breast cancer

12 Isolation of human colon stem cells using surface expression of Protein Tyrosine Kinase Receptor 7 (PTK7)

13 Systems-wide prediction of gene promiscuity reveals a new underground alternative route for 5'-phosphate production in *E. coli*

Figure 4: Overview of the Thesis content. Articles in the main text and in the appendix are shown. The levels of complexity that are explored in each of the studies are highlighted in the corresponding circles (see Figure 3). Note that Article 12 and Article 13 are not put into context because they are the result of minor collaborations, somewhat outside the scope of the global project.

OBJECTIVES

The aim of the Thesis is to explore new ways to connect distinct data types in the field of drug discovery. More specifically:

1. To quantify the clinical importance of biological and chemical traits of drugs, i.e. to assess both the classical chemo-centric view of drug action and the modern approach focused on molecular biology.

   - Article 1 correlates drug targets and off-targets with side effects, and looks for liable chemotypes.
   - Article 2 evaluates the feasibility of the chemo-centric view of diseases, given the available chemical data.

2. To reconcile classical and modern pharmacology by pushing molecular and systems data into e.g. cell line panels.

   - Article 3 evaluates the clinical relevance of drug response measurements in cancer cell lines.
   - Article 4 uses genome-scale metabolic models to better characterize the cell lines.

3. To enable systems pharmacology by bringing objective (2) down to the atomistic level, i.e. to perform structural analysis on a large scale and exert systems-wide perturbations of biological networks.

   - Article 5 extensively comments on the importance of protein structures in the systems era.
   - Article 6 automatizes the identification of protein conformers.
   - Article 7 uses protein structures to propose multi-target strategies.

Part II

ARTICLES

3

ANALYSIS OF CHEMICAL AND BIOLOGICAL FEATURES YIELDS MECHANISTIC INSIGHTS INTO DRUG SIDE EFFECTS

ARTICLE 1   Chemistry, biology and side effects.

AUTHORS   Miquel Duran-Frigola and Patrick Aloy.

JOURNAL   Chemistry & Biology.

TYPE   Research article.

STAGE   Published.

CONTEXT   To establish direct links between clinical data and molecular features, we did an exploratory, agnostic analysis of the importance of drug chemical and biological traits. Drug side effects can be seen as clinical readouts of drug action and here we quantified the relevance of the target- and the chemocentric views.

CITATION   Duran-Frigola and Aloy (2013) [69].

NOTE   Full supplementary material is not provided in this Thesis. In particular, Tables S1 and S2 can be accessed at `http://dx.doi.org/10.1016/j.chembiol.2013.03.017`.

A CHEMO-CENTRIC VIEW OF HUMAN HEALTH AND DISEASE

ARTICLE 2   The chemo-centric view.

AUTHORS   Miquel Duran-Frigola, David Rossell and Patrick Aloy.

JOURNAL   Nature Communications.

TYPE   Research article.

STAGE   Published.

CONTEXT   Given the remarkable predictive power of chemical features observed in Article 1, we decided to get the most out of the chemo-centric view by expanding the chemical space with other small molecules such as environmental substances.

CITATION   Duran-Frigola et al. (2014) [125].

NOTE   Full supplementary data can be accessed at `http://dx.doi.org/10.1038/ncomms6676`.

5

DRUG SENSITIVITY IN CANCER CELL LINES IS NOT TISSUE-SPECIFIC

ARTICLE 3   Drug sensitivity in cancer cell lines.

AUTHORS   Samira Jaeger*, Miquel Duran-Frigola* and Patrick Aloy (*equal contributions).

JOURNAL   Molecular cancer.

TYPE   Research comment.

STAGE   Published.

CONTEXT   Cell-based assays are the modern analogs of the classical phenotype-based approach to pharmacology. Article 2 highlighted the relevance of keeping chemical details, and these are the only molecular information available in standard cell line panels. In this small study, we checked whether, in turn, cell line measurements can be translated to phenotype.

CITATION   Jaeger et al. (2015) [178].

NOTE   This article contains supplementary information that can be accessed at `http://dx.doi.org/10.1186/s12943-015-0312-6`.

6

HARNESSING CANCER'S PH-VULNERABILITY TO IDENTIFY AN OPTIMAL TWO-STEP THERAPEUTIC STRATEGY

ARTICLE 4   Cancer and pH.

AUTHORS   Erez Persi*, Miquel Duran-Frigola*, Patrick Aloy and Eytan Ruppin (*equal contributions; author order to be decided).

TYPE   Research article.

STAGE   In preparation.

CONTEXT   In Article 3, the difficulty to match tissue-specific data with the corresponding cells suggested that molecular characterization is important to profit from cell line panels. We used genome-scale metabolic models to provide such a molecular characterization. On top of these models, to further link the network level with the protein level, we applied pH-activity profile constraints based on sequence data and enzymology assays.

NOTE   This article is awaiting experimental validation.

6.1   ABSTRACT

The initiation and development of cancer is associated with major metabolic alterations. An important aspect of cancer metabolism is the acidification of its extracellular environment and the concomitant alkalization of the cytoplasm, generating a reverse pH-gradient. Although much effort has been devoted to studying the consequences of extracellular acidification of cancer's microenvironment, the role and importance of intracellular alkalization remains poorly understood. Here we provide for the first time a systems biology comprehensive understanding of how changes in intracellular pH (pHi) are coupled to network-wide cancer metabolic alterations, by integrating enzymatic pH-dependent activity profiles into human genome-scale metabolic models of cancer and normal cells. We show that lowering pHi renders cancer cells vulnerable for disruption and contributes to reversing its 'Warburg' nature. This vulnerability is further exploited to identify optimal metabolic targets whose inhibition selectively kills cancer at low pHi. These results unravel an unprecedented role of pHi in cancer metabolism and put forward ground for novel combinatorial efficient therapy.

6.2   INTRODUCTION

Most cancer cells demonstrate prevalent metabolic adaptations, notably the Warburg effect [201], characterized by upregulated glycolysis and lactate production even under aerobic conditions, and an adaptation to low-oxygen (hypoxic) and low-nutrients environments [202]. Another outstanding hallmark of cancer is the acidification of its extracellular environment and the concomitant alkalization of its cell's plasma. This leads to a reverse cancer pH-gradient (pHi $\geqslant 7.2$, pHe $\sim 6.7 - 7.1$) compared with normal cells (pHi ~ 7.2, pHe ~ 7.4) [203], owing to altered activity of various plasma membrane transporters involved in pH regulation and homeostasis [204]. In cancer, these notably include increased expression and activity of acid extruders, such as monocarboxylate transporter (MCT) and Na^+-H^+ exchanger 1 (NHE1), as well as carbonic anhydrases (CAs), which maintain the higher pHi and lower pHe of tumor cells.

This notable reverse pH-gradient is significantly correlated with the spatial gradient of oxygen availability in tumors [205, 206], and is of considerable functional significance. It promotes tumor proliferation, invasion and metastasis [207, 208, 209], as well as aggressiveness and resistance to treatment [203, 210, 202]. Underlying mechanistic explanations have primarily focused on various effects of extracellular acidosis, including the induction of growth factors (e.g. VEGF, HIF), the conversion of secreted lactate to glucose as a nutrient source, suppression of the immune system, and the evolutionary advantage of the surviving cancer cells, which are selected to cope with acidosis, over normal cells in the tumor microenvironment [211, 212].

"...we investigate the effects of interfering with pHi on the metabolic state of cancer vs. normal cells using genome-scale metabolic models..."

Interfering with pH regulation in cancer by inhibition of membrane transporters has been suggested as promising therapeutic potential already in early studies [213, 214]. Some membrane transport targets are today in advanced clinical trials [215]. Such strategy is intrinsically selective (i.e. does not damage healthy cells) and is expected to counteract the effects associated with extracellular acidosis, to some extent. Moreover, it was suggested that such an approach may induce intracellular acidosis which could be toxic to cancer [206], and that the cancer's alkaline intracellular environment is essential to its survival [216].

However, a more comprehensive picture of how changes of intracellular pH (pHi) are coupled to cancer metabolic adaptations and cancer proliferation overall is still lacking. Such an understanding may further facilitate the design of an optimal combined strategy to target cancer vulnerabilities which are exposed at a specific pHi. The difficulty in exploring intracellular effects is partially because of the lack of adequate systems biology methodologies. To this end, here we investigate the effects of interfering with pHi on the metabolic state of cancer vs. normal cells using genome-scale metabolic models (GSMM) of a wide array of cancer and normal cell-lines. To account for the effect of pHi we developed a novel methodology which predicts the pH-dependent activity profiles of enzymes and then integrates the latter into GSMMs. We identified an

unprecedented role of pHi regulation in cancer metabolic adaptations and survival capacity. We show that, compared to healthy cells, cancer cells are more adapted towards alkaline high-pHi and more vulnerable in the acidic low-pHi regime. Moreover, at high-pHi cancer develops higher glycolysis and adapts to hypoxia. Similarly, at low-pHi these metabolic adaptations are reversed. Furthermore, we identify key enzymes whose activity strongly affects cancer vulnerabilities at low-pHi. Building on this finding, we predict new drug targets that are both selective (inhibit cancer cells primarily) and pH-specific (inhibit cancer proliferation more effectively at low-pHi than at high-pHi). These findings form a basis for a combined therapeutic strategy targeting cancer.

6.3 RESULTS AND DISCUSSION

To investigate the effects of pH on the metabolic activity of cells we performed a two-step computational analysis. First, we developed and validated a new framework that predicts the activity of each metabolic enzyme at a wide pH range. Second, we integrated these predictions to evaluate the activity of normal and cancer metabolism at the network level, using genome scale metabolic modeling (GSMM).

pH profiles of metabolic enzymes were fetched from *in vitro* measurements available in the BRENDA database, complemented with homology-based predictions (see Section 6.4 and Section 6.5, Figure 41, Figure 42, Figure 43 and Figure 44). As shown in Figure 38A, a pH-activity profile here is defined by the acidic and basic critical points corresponding to 0%, 50% and 100% of activity. When no experimental value was available, we could predict it based on homologs of the enzyme of interest, and the values of the other critical points. Predicted vs experimental optimal pH values showed a Pearson's correlation of 0.759 in a 10-fold cross-validation (Figure 38B); this marked correlation held for the rest of critical points (see Section 6.5, Figure 45, Figure 46 and Figure 47). Reassuringly, the predicted optimal pH of the enzymes were remarkably consistent with the known pH range of the cellular compartments in which they reside (Figure 38D) [217, 204]. Lysosome and, to a lesser extent, Golgi apparatus are at acidic physiological pH: accordingly, enzymes in these compartments had low optimal pH values. On the contrary, we obtained basic optima for enzymes operating in the mitochondrion and peroxisome, in good agreement with the characteristic pH of these organelles.

"...the predicted optimal pH of the enzymes were remarkably consistent with the known pH range of the cellular compartments in which they reside."

To investigate the effects of an induced pHi regime (which can be achieved in practice by inhibition of known membrane transporters, e.g. [218]) on the metabolic behavior of cancer and healthy cells, we integrated the pHi profiles computed above for every metabolic enzyme into genome scale models of normal and cancer metabolism. This is achieved by modifying the bounds on the flux of each enzyme in the network as a function of its predicted activity at a given pH (see Section 6.4). This enables us to compute the flux range carried by any enzyme for each cell modeled at any given pHi, and as a result, the cell's
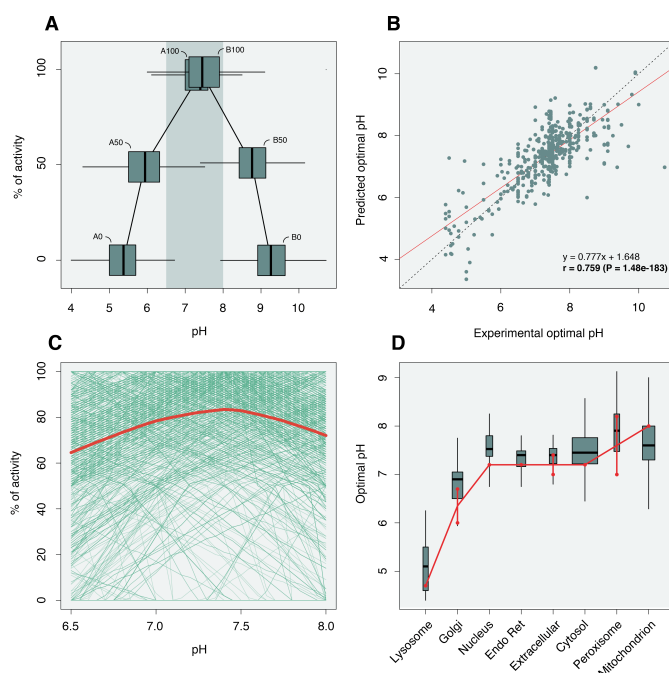
Figure 38: Enzymatic pH-activity profiles. (A) For each enzyme, six critical pH points are defined, namely 'A0', 'A50', 'A100', 'B100', 'B50' and 'B0'. At each point, the distribution of pH values is shown; including experimental as well as predicted values. The region of interest for further GSMM analyses is dark-shaded. (B) Predicted optimal pH vs experimentally observed optima. Optimal pH is defined as the average of 'A100' and 'B100'. The red line corresponds to the linear regression, accompanied with the Pearson's correlation coefficient (r). (C) Predicted pH-activity profiles in the range of $6.5 - 8.0$ pH units. Each of the green lines corresponds to a pH-activity profile of a single enzyme. The red line represents the average pH profile. (D) Optimal pH values of metabolic enzymes in each cellular compartment. Box widths are proportional to the number of enzymes in each category. Red dots delimit the physiological pH-range of the compartment.

proliferation rate and various other metabolic capacities at any given pHi level. We performed this integration and modeling in a series of cell-specific models of the NCI-60 cell-line panel and of the HapMap healthy lymphoblastic panel that we have just published and validated [219, 33].

The resulting effect of pHi on cellular proliferation in both normal and cancer cell types is shown in Figure 39. As evident, cancer cell growth rate is lower at the acidic low-pHi regime than normal cell growth, but the situation markedly reverses at alkaline high-pHi levels, where the growth of cancer cells is higher and much more robust to pHi alterations than the growth of normal, healthy cells. Notably, this effect is robust with respect to noise in the pH-profile values, and reassuringly vanishes under random (i.e., wrong) assignment of pH-profiles to enzymes (Figure 48). Furthermore, trying to understand which enzymes are essential for this behavior we identified, by divide and concur

search, that in the glycolysis pathway, the moonlighting enzyme GAPDH (and its paralogue GAPDHS), strongly affect cancer's pH-dependent behavior. Their inhibition arrests cancer's proliferation at pHi levels below the physiological range, exposing their low-pHi vulnerability and amplifying the apparent selective effect between cancer and normal cells.
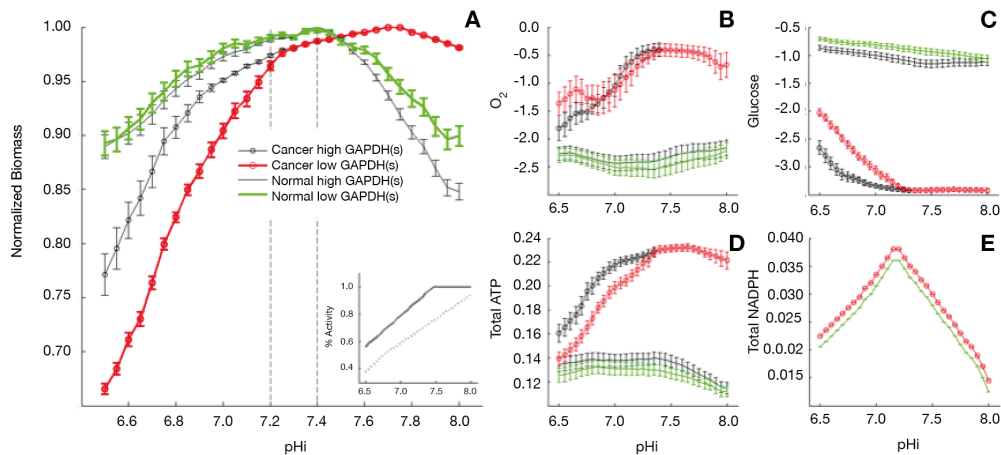
Figure 39: Key metabolic characteristics of cancer (circles) and normal healthy (solid curves) cells as a function of pHi levels. Results are displayed twice: when GAPDH and GAPDHs carry high flux (black) and when they carry low flux (color), as indicated by the inset of subfigure A. (A) Cellular proliferation (biomass production yield) as function of pHi, normalized by the maximal value obtained across all pHi levels examined. Dashed lines indicate the physiological range of pHi in both cells types ($\sim 7.2 - 7.4$). (B) Oxygen consumption rates. (C) Glucose uptake rates. (D) Total ATP production rates (E) Total NADPH production rates. Rates are given in mmol/mgDW/h. Note that uptake rates (B-C) are conventionally depicted with a negative sign, so that lower (more negative) values denote higher uptake rates. Mean values and their corresponding standard deviations are estimated across two populations of GSMMs representing the NCI-60 (cancer) and HapMap (normal lymphoblastic) cell-lines panels (Methods). These results are robust with respect to choice of parameters (Figure 49).

"... the models capture the more glycolytic and hypoxic nature of cancer cells.

The effects of different pHi levels on oxygen and glucose consumption rates (Figure 39B-C) and on total ATP and NADPH production rates in cancer and normal cells are also shown (Figure 39D-E). Overall the models capture the more glycolytic and hypoxic nature of cancer cells. Evidently, oxygen consumption rate is compromised in the high-pHi regime while glucose uptake rate increases, indicating that the alkaline intracellular pH indeed contributes to the hypoxic and high-glucose characteristics of cancer cells. ATP production is optimal at physiological pHi and robust at high-pHi. All of these effects are reversed to some extent in the acidic low-pHi regime. Hence, lowering pHi is predicted to render cancer cells to a less glycolytic and a more normoxic metabolic state. The effect of reversing cancer's glycolytic nature, and hampering its ATP production capacity are further accentuated when GAPDH/S are inhibited. Notably, NADPH (as well as NADP production) is optimal at physiological pHi

and sensitive to any change in pHi, affecting both cell types similarly, in a non-selective manner.

Taken together, these results testify that the specific resilience of cancer cells to the alkalization of their intra cellular environment is a prime driving force that enables them to generate and maintain the acidic extracellular gradient to which so many potential cancer-beneficial functional effects have already been attributed to. As shown, in contrast to healthy cells, cancer proliferation is hardly diminished at moderate alkaline pHi levels, which provide cancer cells with specific selective proliferation advantages versus their surroundings. This suggests that acidifying perturbations aiming at lowering the pHi will in turn increase the vulnerability of cancer cells, and, as evident from Figure 39, overall tend to have a larger detrimental effect on cancer than on normal cells proliferation. Such an effect can be further amplified by partial inhibition of GAPDH/S enzymatic activity. This view thus suggests a two pronged therapeutic attack on cancer, first by acidification agents, followed by drugs that exploit the specific vulnerabilities exposed at the latter state.

"To identify potential post-acidification metabolic anticancer targets we simulated the inactivation (knockout) of each metabolic gene…"

We hence turned to simulate the metabolic acidic state in cancer cells and identify pH-specific selective targets. To identify potential post-acidification metabolic anticancer targets we simulated the inactivation (knockout) of each metabolic gene by constraining the flux of the reactions affected by it (see Section 6.4) in two pHi regimes: 'physiological' (i.e., pHi = 7.3) and 'low' (i.e., pHi = 6.7). We explored the effect of gene inactivation on cellular proliferation and on the consumption and production rates of key metabolites across the cancer and normal cell-lines. Targets of particular interest are those that are (i) *Selective*: that is, hamper proliferation of cancer cells at low-pHi selectively, i.e., with a lesser inhibitory effect on the proliferation of normal cells, and (ii) *pH-specific*: that is, hamper cancer's proliferation at low-pHi more than they do at physiological-pHi. This ensures that the selective killing of cancer by inhibiting these targets is not exacerbated at low pHi, but rather amplified. To identify such targets we introduce two corresponding scores: selectivity and pH-specificity, to evaluate and compare gene knockouts (see Section 6.4). The higher these scores are the more selective and pH-specific are the predicted targets in killing cancer cells effectively at low-pHi.

Figure 40 presents a classification of the gene targets, their gene-inhibition scores and the impact of their inhibition on key metabolites relative to wild type (WT). The inhibition of 12 genes was found to be both selective and pH-specific in all cancer-normal pair comparisons. Additional 11 targets were found to be selective with low pH-specificity scores. Moreover, some of these targets further contribute to the increase in oxygen consumption rate and concomitantly to the decrease in glucose consumption, in cancer at low pHi.

The predicted targets include a number of genes in the glycolysis and pentose-phosphate-pathway. An interesting prediction is the glycolytic enzyme GPI,

whose scores are positive and its inhibition decreases glucose consumption and increases oxygen consumption more than other targets in cancer cells at low pHi. Hence, GPI is predicted to significantly contributing to further killing of cancer at low pHi and reversing cancer's Warburg characteristics at this pHi.
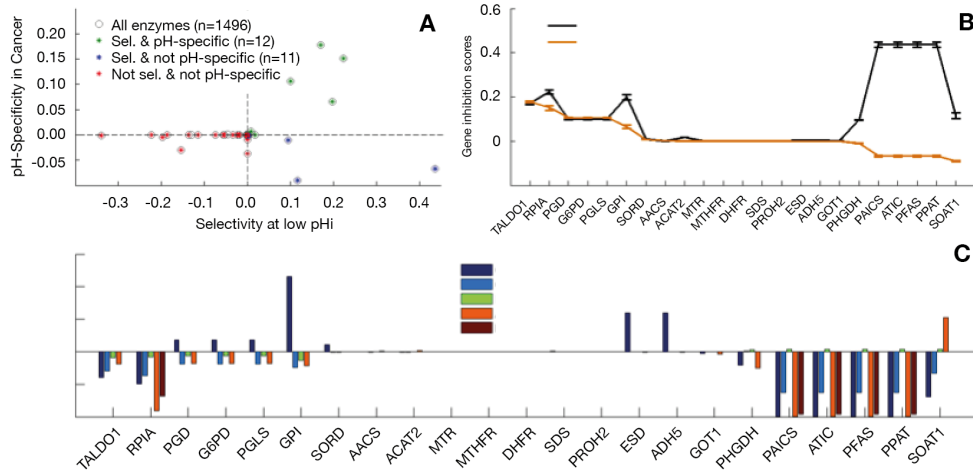
Figure 40: Gene Inhibition (Knockout) Analysis. (A) Clustering of all gene knockouts according to their selectivity and pH-specificity scores. Positive scores identify 12 selective and pH-specific targets (green). Additional 11 targets are identified as selective but not pH-specific. These results are robust with respect to level of gene inhibition, the exact choice of low pHi, and the constraint cell's proliferation (Section 6.4 and Figure 50). (B) The predicted selective and/or pH-specific 23 targets that recur in all cancer-normal comparisons, ranked by their average pH-specificity. The average selectivity score is superimposed. (C) The effect of knockdown of the identified 23 targets on other key metabolite consumption/secretion relative to the wild type (WT). Average changes across all cancer cells at low pHi are shown.

Overall, reviewing our results we rigorously show that on the system level, lowering pHi may be combined with the inactivation of selective and pH-specific gene targets to put forward an optimal two-step therapeutic strategy. First, lowering pHi to hamper cancer's proliferation and channel it to a less glycolytic and more normoxic metabolic state. Second, further exploit these modifications by inactivating selective and pH-specific targets at this low-pHi to both further decrease cancer proliferation and further reverse cancer's Warburg nature.

6.4 METHODS

6.4.1 *Construction of pH-activity profiles*

Dependency of enzymatic activity on pH was obtained from experimental data in BRENDA (June 2014) [brenda-enzymes.org]. As shown in Figure 38A, for each enzyme we aimed at six critical points, corresponding to the lower and upper limits of 0%, 50%, and 100% of activity. Experimental points of 100%

activity were mainly obtained from the 'pH Optimum' field in BRENDA. 0% and 50% points were read from the 'pH Range' category, after manual curation: records reporting activities up to 25% were approximated to 0%; activities from 25 to 75% were set to 50%; and activities above 75% were set to 100%. When more than one record was available, we extracted the median value.

The vast majority of experimental values corresponded to 100% of activity, i.e. the optimal pH (Figure 41). We learned linear regressors based on experimental data and values of close homologs in order to impute missing critical points. Then, we screened the enzymes in the GSMM against the pH-profile database using JackHMMER, obtaining profiles for $1,444$ of the $1,905$ metabolic genes (76%). A detailed description of the method is provided in Section 6.5.

Once critical pH points were obtained, the % of activity of an enzyme at any given pH could be linearly interpolated. Metabolic genes without a predicted profile were conservatively given a constant activity of 100%, such that no constraints are applied to them in the genome-scale metabolic models.

6.4.2 *Genome-scale metabolic modeling and application of pH-activity profiles*

We applied pH-activity profiles into a panel of genome-scale metabolic models (GSMM), developed recently [219] based on the human GSMM [29], representing the NCI-60 cell-lines panel and the normal healthy lymphoblastic cell-lines from the HapMap project. This panel of models capture key differences between cancer and normal cells, including Warburg characteristics. Moreover, they have identical sizes and are modeled under identical media composition (RPMI-1640), hence highly adequate for comparative analysis.

Constraint-based modeling of metabolic network

Constraint-based Modeling (CBM) approach imposes mass-balance, thermodynamic and enzymatic capacities constraints to define the allowable functional states of biochemical genome-scale model [220]. These constraints can be mathematically represented as linear equations:

$$\frac{dx}{dt} = S \cdot v = 0 \tag{2}$$

$$v_{min} \leqslant v \leqslant v_{max} \tag{3}$$

where v is the network's flux vector and S is the $m \times n$ stoichiometric matrix, where m and n are the number of metabolites and reactions, respectively. The matrix specifies all biochemical reactions and metabolites in the network. Constraint Equation 2 assures steady state, where the production and consumption

rate is equal for each metabolite in the network. Thermodynamic and enzymatic capacities are defined by imposing bounds on the reactions' fluxes and are embedded in constraint Equation 3. The permissible flux that a reaction cay carry, in a given metabolic state, is estimated using flux-balance-analysis (FBA), and flux-variability-analysis (FVA) [31], taking its maximal flux as a proxy for its catalytic activity.

Cellular proliferation is therefore estimated as the maximal flux carried by the biomass reaction in the GSMM, which represents the cellular growth yield. Since the cells we model are highly proliferative, we constraint cellular proliferation to be larger than X% of its maximum in order to infer the activity of all other reactions. The results reported here are robust with respect to the choice of X in the range $X = 70\% - 90\%$ (Figure 49).

Integration of pH-activity profiles

We apply the pH-profiles of Figure 38 to adjust the bounds of each reaction in the GSMM, at a given pH. This is done in three steps. First, in a given pH, the activity of genes relative to their maximal activity (at their optimal pH) defines a pH-specific activity of each gene, $W_G = [0, 1]$. Second, taking W_G of all genes, we infer the activity of the reactions in the GSMM, based on the embedded genes-reactions logical rules, associated with each reaction. For an 'AND' logic the minimal W_G is assumed, and for an 'OR' logic the maximal W_G is assumed. Hence, we end up with a weight factor, $W_R = [0, 1]$, for a reaction R. Third, the upper and lower bounds of a reaction are scaled by W_R. For a bi-directional reactions ($v_{max} \geqslant 0$ and $v_{min} \leqslant 0$) the new lower bound is $LB = W_R \times v_{min}$ and the new upper bound is $UB = W_R \times v_{max}$. To avoid invalid ranges of bounds, for a forward reaction ($v_{max} > v_{min} \geqslant 0$) only the upper bound is scaled by W_R, ensuring that $UB \geqslant v_{min}$. Similarly, for a reverse reaction ($v_{min} < v_{max} \leqslant 0$) only the lower bound is scaled by W_R, ensuring that $LB \leqslant v_{max}$. We assume that cellular organelles are well buffered and therefore apply these modifications only to cytosolic enzymes.

Gene knockout simulation and analysis

The knockout (KO) of a gene G is simulated by setting W_G to $0 - 0.1$, representing activity inhibition of $100\% - 90\%$ respectively. The effect of a gene KO on cellular proliferation is estimated by $nB_{KO,pH}^{Cancer} = B_{KO,pH}/B_{WT,pH}$, where $B_{WT,pH}$ is the biomass of the wild-type (WT) at a given pH, and $B_{KO,pH}$ is the biomass of the cell following gene KO at this pH. To assess the importance of a gene KO, we introduce two ranking measures:

1. *Selectivity score* (SEL), which measure the difference in cellular proliferation between cancer and normal cells following gene KO. Hence, $SEL = nB_{KO,pH}^{Normal} - nB_{KO,pH}^{Cancer}$. The larger SEL the more selective is the gene KO.

2. *pH-specificity score* (PHS), which measures for a given cell (i.e., normal or cancer) how effective is a gene KO at 'low' pH (pH = 6.7) compared with

its effect at 'physiological' pH (pH = 7.3). Hence, $PHS = nB_{KO,pH=7.3} - nB_{KO,pH=6.7}$. The larger PHS the higher is the effect of the gene KO at low-pH compared with its effect at the higher pH.

The ranking of genes by either SEL or PHS is highly robust within the gene inhibition range $W_G = [0, 0.1]$, and is insensitive to the exact choice of 'low' pH (Figure 50).

6.5 SUPPORTING INFORMATION

6.5.1 *Predicting pH-activity profiles of human metabolic enzymes*

Background

Tight regulation of intracellular pH is a prominent principle of living systems. Such regulation is needed for at least two reasons. First, because cells fundamentally transduce energy through proton gradients and proton-coupled electron transfer reactions. Second, because pH determines the charge state of weak acids and bases, affecting physical and physiological properties of biomolecules such as proteins [221].

Proteins need to be stable at the characteristic pH of the subcellular environment in which they operate. Protein pH-stability is mainly driven by amino-acid composition and 3D disposition of titratable groups. In the recent years, several computational methods have been developed to predict the pH of optimal stability of proteins, requiring e.g. the calculation of residue pKas in folded and unfolded states, or the proportion of acid and basic residues in buried regions [222].

However, before stability, pH fluctuations may directly affect protein function. Enzymes, for instance, have evolved to perform efficient chemical reactions involving electron transfers and proton translocations. Accordingly, there is a trade-off between reactivity and stability in the active site [223]. Besides fold denaturation, loss of enzymatic activity at aberrant pHs can be thus ascribed to protonation states of catalytic residues, slight spatial modifications that hamper substrate or cofactor binding, or differential protonation of the substrate itself. Unlike pH-stability calculations, no physics-based method is yet able to integrate these factors and infer the one that limits activity, which many times depends on subtle phenomena, and predicting pH-activity profiles (i.e. the full curve that relates enzymatic activity to pH) is a formidable task. These curves are usually bell-shaped around an optimal pH, although they can be also asymmetrical and, occasionally, multiple optima can be observed.

Fortunately, for decades enzymes have been the matter of thorough biochemical study. A key parameter in enzymology is the pH of the buffer solution, which needs to be optimized in order to identify the appropriate conditions for the assay. As a result of this process, pH-activity curves, or at least some critical points, are reported in many scientific publications. The BRENDA database [brenda-enzymes.org] is devoted to compiling this and other information from the literature. Nowadays, full or partial information on the pH sensitivity of enzymes can be found for thousands of enzymatic functions in thousands of organisms.

The corpus of pH-activity information available in BRENDA allows us to explore the hypothesis that homologous enzymes will respond similarly to changes in pH. Homologous proteins are similar in sequence and fold, and many times catalyze analo-

gous reactions. Therefore, it is reasonable to postulate that pH fluctuations will affect homologs to a comparable extent, yielding similar pH-activity curves. In this study, based on the experimental data available in BRENDA, we embraced this hypothesis in the context of human metabolic enzymes. Concretely, we aimed at the prediction of pH-activity profiles of enzymes contained in the Recon1 reconstruction of human metabolism [29].

Experimental data

RAW PH-ACTIVITY DATA   In order to compile an experimental dataset, we fetched all 'pH Optimum' and 'pH Range' records from BRENDA (July 2014). These included a total of $54,462$ entries, $34,493$ of them having comments edited by BRENDA curators.

CURATION OF PH RECORDS   In particular, 'pH Range' entries had comments on the extent of activity loss. These were the ones that required, most often, manual curation. Percentages of activity loss are not specified in BRENDA and were extracted from these comments. Whenever no percentage value was provided, we assigned a 50% activity loss to descriptions referring to 'half activity', 'marked reduction', or similar, and 0% activity to cases like 'complete loss of activity' or 'no activity'. Curation of the optima was less laborious, although sometimes we found disagreement between the comments themselves and the values presented by BRENDA. We disambiguated such cases by taking the value reported in the description.

DISCRETIZATION OF CRITICAL PHS   To delineate activity profiles that were suitable for machine learning, we defined six critical pH points, namely 'A0', 'A50', 'A100', 'B100', 'B50', and 'B0'. These correspond to the acidic (A) and basic (B) limits of 0%, 50%, and 100% of activity. Accordingly, activity values below 25% were rounded to 0%, and values between 25% and 75% were approximated to 50%. Activities above the 75% were set to 100% if no optima were available. Regarding the 100%-activity points, when a single optimum was reported —as it was most commonly the case—, we assigned the same value to 'A100' and 'B100'. Often, BRENDA contained several records per enzyme; to minimize the impact of outliers, we calculated the median in these cases. If more than one optimum were explicitly reported, we defined a broad peak limited by the interquartile range.

In Figure 41 we present global statistics on the database. In total, we collected $23,658$ enzyme entries, corresponding to $4,766$ functions (4th EC level) in $4,521$ organisms. Notably, 926 records corresponded to human enzymes, and $3,902$ to mammals. As it can be seen in Figure 41A, for the vast majority of cases we had information on optimal pHs, while half-activity points were less frequent, and complete loss of activity points were rare. Figure 41B shows that these experimental values spanned a wide range of pHs.

SEQUENCE ANNOTATION   Additionally, we assigned sequences to as many BRENDA enzymes as possible. Often, BRENDA itself provided sequence annotations and, in other cases, this information was extracted from UniProt if the BRENDA organism could be mapped to NCBI taxonomy. We completed most of the remaining entries by collecting proteomes at the species level from UniProt, and running ADIOS on these proteomes to assign EC numbers [224]. Similarly, the remaining cases were obtained by building HMM profiles across the full EC annotation in UniProt [225], and then running HMMSearch with default parameters on the proteomes (please note that this annotation procedure is analogous to that in BRENDA, where sequences are often annotated based on BLAST results). As it can be seen in Figure 41C, we could as-
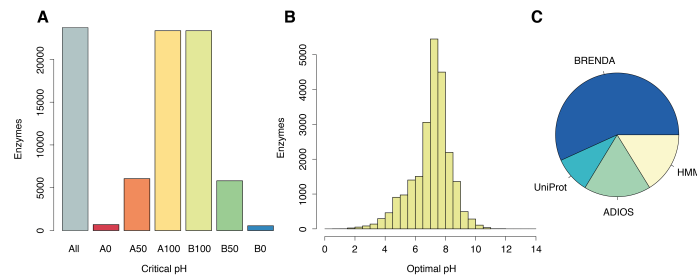
Figure 41: Information extracted from BRENDA. (A) Number of enzymes for which experimental data were available from BRENDA, at each critical point (acidic and basic limits of 0%, 50% and 100% of activity). For most enzymes, the 100%-activity points were available. (B) Distribution of optimal pHs (average between 'A100' and 'B100') across all enzymes. (C) Sources of sequence annotation, being BRENDA annotation priority, followed by UniProt, then ADIOS, and finally HMM search.

sign sequences to most enzymes, coming two thirds of them from BRENDA itself or UniProt.

Proof of principle

HOMOLOGOUS ENZYMES HAVE SIMILAR PH VALUES   As explained before, we work on the assumption that homologous proteins will yield similar pH-activity curves. Before proceeding, we checked whether this is actually the case. As it can be seen in Figure 42, enzymes performing the same function (same EC number) have similar pH optima ('A100', 'B100') and their activities decay equally as we deviate from these optima ('A50', 'A0', 'B50', 'B0'). Note, also, that best precision could be achieved for 100% activity points.

In the database, as expected, we observed a very strong correlation between function and sequence homology: setting a homology threshold of E-value 1×10^{-4}, we observed a strong enrichment of homologous sequences among enzymes with the same EC number (odds ratio of a right-tailed Fisher's exact test of 995.3, P-value ~ 0). Overall, this suggests that sequence, function and pH profiles are tightly related.

PREDICTED PHS OF OPTIMAL STABILITY DO NOT CORRELATE WITH EXPERIMENTAL ACTIVITY OPTIMA   An alternative approach to ours would be to use pH-stability curves to estimate pH-activity profiles. Some authors suggested that certain degree of correlation exists between the optima of stability and activity [223]. Following this notion, we used the well-established PROPKA tool (v3.1) to calculate the pH-stability curves based on 3D structures. In Figure 43A, we show how these stability curves, based on pKa titration of ionizable residues, are difficult to match with experimental optima of activity. In general, we could not find a clear correlation between calculated optimal stabilities and pHs of highest activity (Figure 43B). Moreover, in our hands, stability curves were not useful to predict 50% and 0% critical points.

Database of pH-profiles

Given that, to date, no physics-based method is able to systematically predict pH-activity curves, we decided to only capitalize on the observation that homologous sequences have, in general, similar pH-profiles. The method that we developed for this
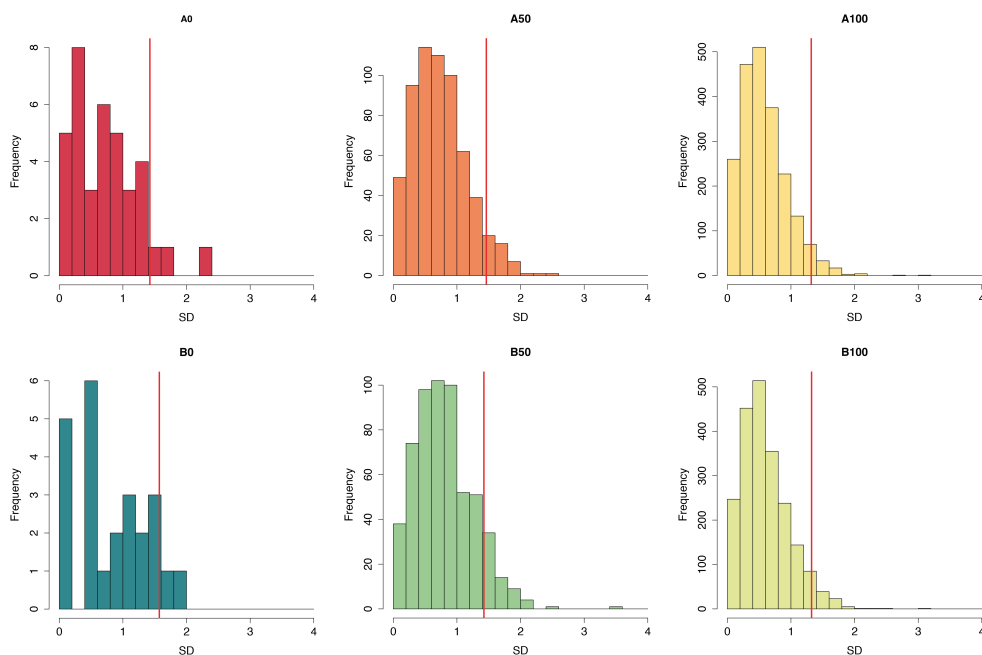
Figure 42: Standard deviation (SD) of critical pH values within the same EC number (only ECs with at least three instances included). The red line denotes the background SD, i.e. the global SD across all enzymes discarding EC information.
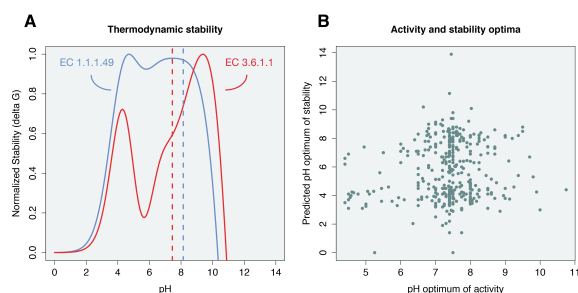
Figure 43: Lack of correlation between predicted pH of optimal stability and activity. (A) Stability curves of two human enzymes for which structures were available. Dotted lines denote the experimental optima reported in BRENDA. (B) Calculated optima of stability vs optima of activity.

means is schemed in Figure 44. In brief, we filled the gaps of the experimental database in order to have a more complete database of profiles ('A0', 'A50', 'A100', 'B100', 'B50', 'B0') that could be later queried. To this end, we predicted missing critical pH values mainly based on other critical points, and knowledge on homologous enzymes. The net result of this step was a database containing full pH-activity profiles that were a mixture of experimental and predicted values (see details below).

IMPUTATION OF CRITICAL PH VALUES   As evident in Figure 41, most often we did not have all six critical pH values experimentally available for an enzyme. In BRENDA, optimal (100% activity) pHs are usually reported, but 50% and 0% activities are more rare. We developed a method to predict these missing points and ultimately
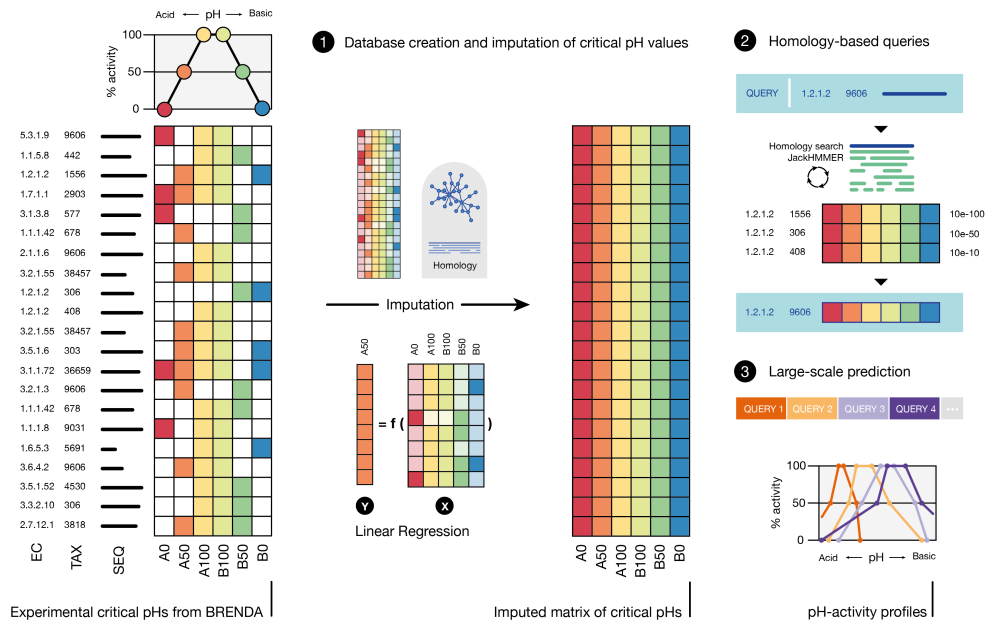
Figure 44: Scheme of the pH-profiling method. (1) We impute the missing values in the experimental database in order to obtain full activity profiles for each enzyme. For this, we first build a preliminary matrix based sequence homology. Then, we use this matrix to build linear regressors, whose predictions are used to impute the experimental matrix and yield a fully filled database of profiles. (2) To predict pH-activity profiles for a new enzyme, we search this database using JackHMMER. (3) This querying procedure can be easily performed on a large scale.

have full profiles for all enzymes in the database. The predictive models were built as follows:

1. We filled a preliminary matrix containing values based on close homologs (rows: enzymes; columns: critical pHs). For this, we previously constructed a network of enzymes in BRENDA where two enzymes were connected if they had sequence relatedness (computed with JackHMMER, with default parameters, E-value cutoff 1×10^{-4}). Then, to complete each row (i.e., each enzyme) we looked for the closest homologs in the network with available experimental data (removing itself from the search), and calculated the weighted median. Here, weights corresponded to the $-\log$ E-value.

2. The preliminary homology-based matrix above was used to avoid empty cells in the training of linear regressors. That is, in order to have a final, refined fully filled matrix, we built a simple linear model for each of the six critical pH points. To predict, for instance, 'A50', we used 'A0', 'A100', 'B100', 'B50', and 'B0' as variables from the preliminary matrix (1).

3. Finally, further fine tuning of the pH values (first or second decimals) may be obtained by up-weighting columns such as 'A100' and 'B100', where experimental data were more abundant and therefore homology-based values more reliable. Concretely, we readjusted those predictions that had a homology-based value within ± 1 pH unit of the predicted one. In this readjustment, we simply weighted by the proportion of experimental data in the column, from no influ-

ence (no weight) of the homology-based value in absence of experimental data, to the simple average with the predicted one in the hypothetic case of a fully experimental column.

INTERNAL VALIDATION OF THE REGRESSIONS   We performed a 10-fold cross-validation of each of the regressors, obtaining R-squared values of $\sim 0.6 - 0.8$, which were slightly improved after the homology-based readjustment.

Querying the pH-profile database

Once the pH-profile database was completed, we screened our human enzymes of interest against it (Figure 44). For this, we mapped genes in Recon1 to UniProt identifiers using UniProt's IdMapping tool. EC numbers were also extracted from UniProt. In total, we could assign EC numbers to $1,444$ of the $1,905$ genes in Recon1. Using the pH-profile database, we tried to infer the pH-profile of these enzymes.

FUNCTIONAL MATCHING   The first step in the search was to look for records having the same EC number than the query enzymes. Please note that, even when we found the exact human record (which was the case for ~ 500 enzymes), this doesn't mean that the full experimental profile was available. Rather, most of the times it meant that the experimental optimum was available, but not the rest of critical points, which were predicted in the imputation step.

HOMOLOGY-BASED WEIGHTING   When several records from different species were matched, and if sequences were available, we performed a JackHMMER search in order to assign E-values of homology. Then, each critical point was calculated as a weighted median; weights corresponding to the $-\log$ E-value (upper limit of 200). Preference was given to enzymes with more abundant experimental data by exclusively selecting them if they had a $-\log$E-value above 100. After this procedure was done, we had, for almost $1,500$ enzymes in Recon1, the acidic and basic limits of none, half, and full activity, i.e. their 'A0', 'A50', 'A100', 'B100', 'B50', and 'B0' critical pHs.

Validation

Before applying these pH-profiles to GSMM, we performed an extensive validation of the method.

10-FOLD CROSS-VALIDATION   We submitted the predictions on the Recon1 enzymes to a 10-fold cross-validation. In order to avoid over-fitting, we split training and test sets from the very initial data, i.e. before the homology-based imputation. In addition, since more than one gene in Recon1 could have the same EC number, we removed all human records in the training set that had EC codes represented in the test set. Results of the 10-fold cross-validation are shown in Figure 45.

REMOVAL OF HUMAN ENZYMES   In addition, we performed a stringent validation by removing all human enzymes from the initial dataset. Results are shown in Figure 46.

REMOVAL OF HUMAN ENZYMES AND EC NUMBER INFORMATION   Finally, a yet more stringent validation was done after, in addition to removing all human enzymes, EC information was excluded from the search. Here, in practice, all predictions were based on the JackHMMER search, without functional (EC) supervision. Figure 47 shows that the method continues to perform well in this case.
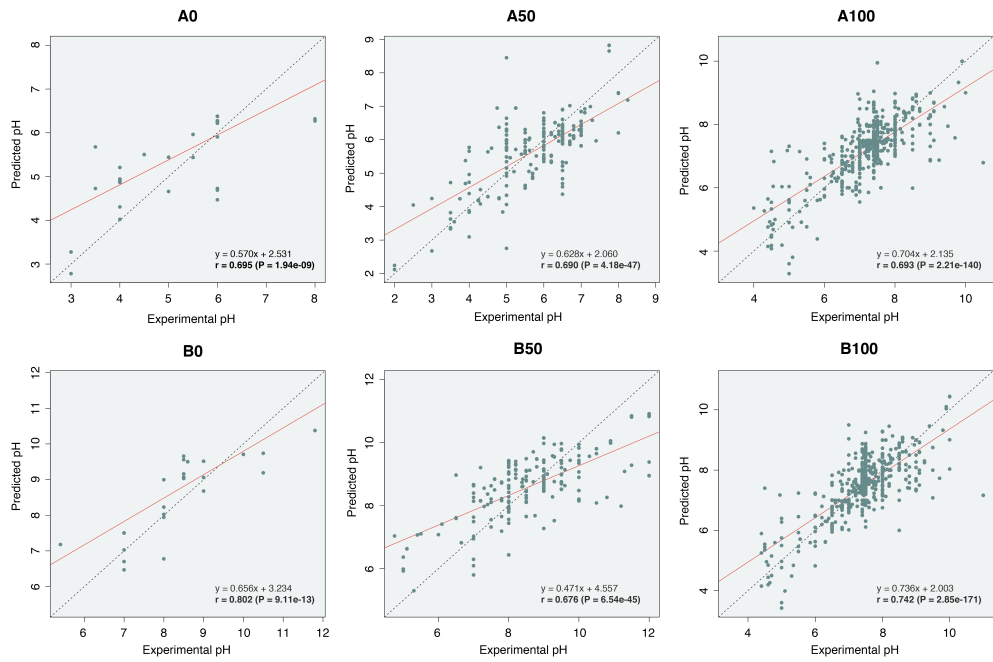
Figure 45: Predicted vs experimental critical pHs for the human dataset, after a 10-fold cross-validation. The points correspond to the predictions on the test sets. Pearson's correlation coefficients (r) are embedded within the plots.
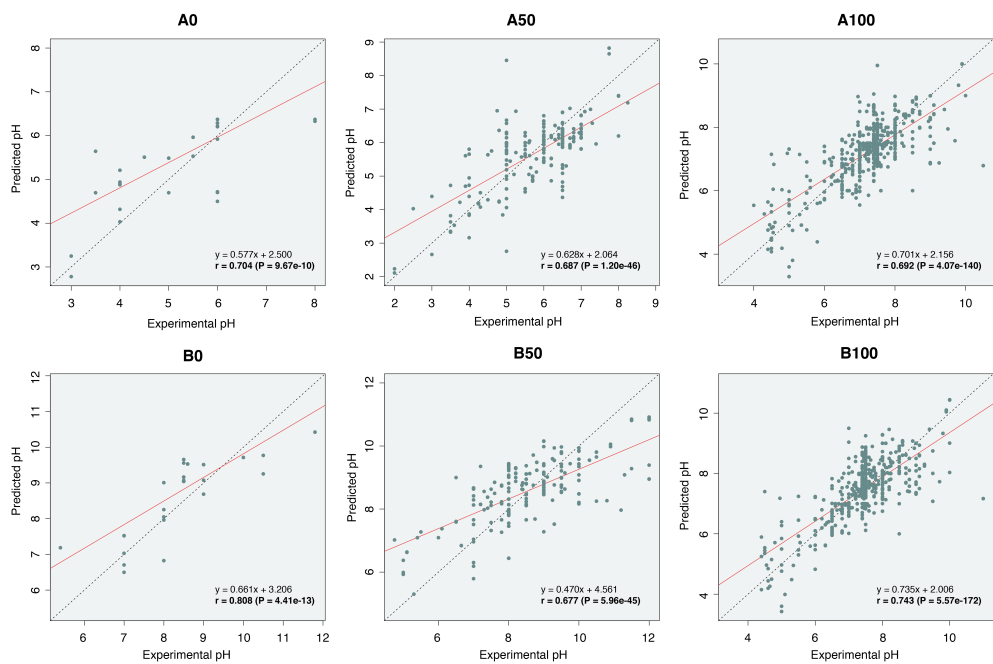
Figure 46: Similar to Figure 45, predicted vs experimental critical pHs, when all human enzymes are removed from the training set.

6.5.2   *Genome-scale metabolic modeling*
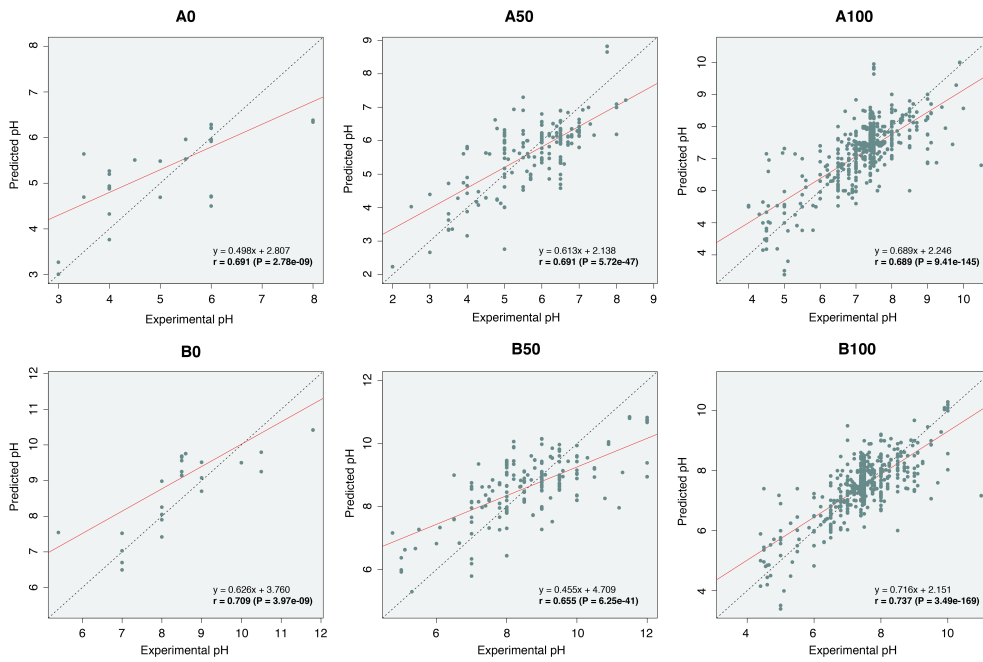
Figure 47: Similar to Figure 45 and Figure 46, predicted vs experimental critical pHs, when all human enzymes are removed from the training set, and EC number information is removed from the test set.
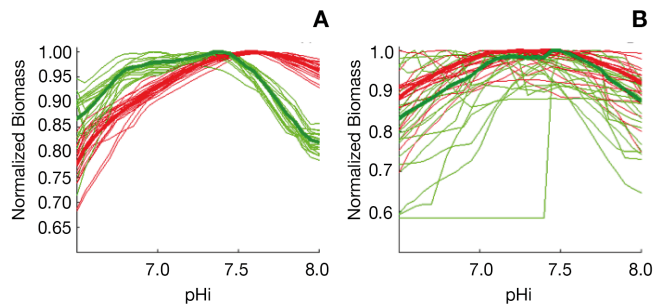
Figure 48: Robustness analysis performed on a single cancer-normal cell-lines pair. (A) the effect of perturbing the enzymes' pH-profiles by Gaussian noise ($\mu = 0$, $\sigma = 0.1$) on the cellular proliferation of a cancer cell (red) and healthy cell (green) as function of pHi. Normalized biomass of 20 independent realizations of the Gaussian noise is shown in solid-thin curves. Solid-thick curves depict the average profiles. (B) the effect of random (i.e., wrong) assignment of pH-profile to enzymes. Differences between cancer (red) and healthy (green) cells are lost, specifically at low-pHi. Solid-thin curves represent 20 independent realizations of random assignments. Solid-thick curves depict the average behavior.
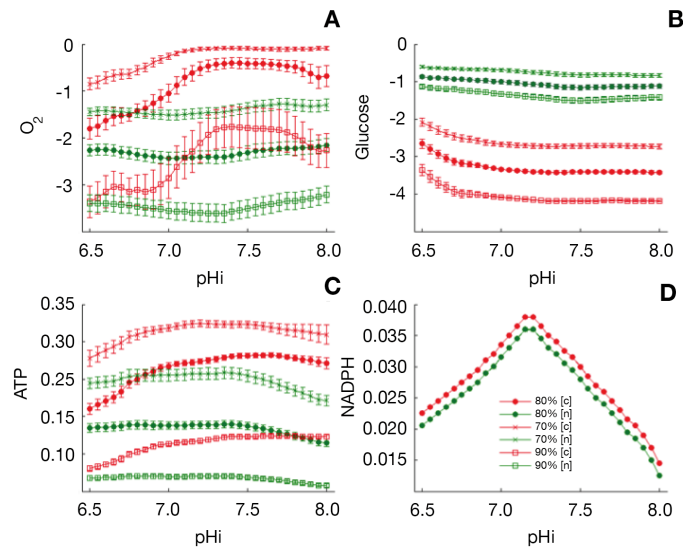
Figure 49: Robustness analysis across the panel of cancer (red) and normal (green) cell-lines examined in Figure 39 of main text, demonstrating that the behavior of key metabolites, i.e. oxygen consumption rates (A), glucose consumption rates (B), ATP production rates (C) and NADPH production rates (D) hold for various choices of the constraint on the objection cellular function, i.e., larger than X% of the FVA biomass maximum (see Section 6.4). Shown are the behaviors in the range X = [70% − 90%]. As depicted, this parameter only scales the absolute values of metabolite consumption/production but does not change the overall trends observed across the pHi and the evident differences between cancer and normal cells.
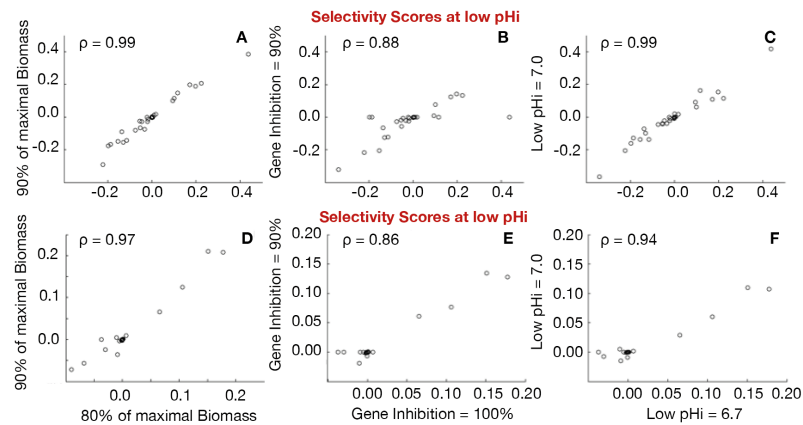
Figure 50: Sensitivity of gene inhibition scores. Selectivity (A-C) and pH-specificity (D-F) to choice of parameters in the genome-scale metabolic models of cancer and normal cells. The correlation between the score is shown for: (A,D) different choices of the constraint on the minimal percentage of the FVA-maximal biomass production [80% − 90%]. (B,E) different choices of the percentage of gene inhibition percentage [90% − 100%]. (C,F) the value of 'low' pHi chosen [6.7 − 7]. ρ values are Pearson's correlation coefficients.

7

STRUCTURAL SYSTEMS PHARMACOLOGY: THE ROLE OF 3D STRUCTURES IN NEXT-GENERATION DRUG DEVELOPMENT

ARTICLE 5   Structural systems pharmacology

AUTHORS   Miquel Duran-Frigola*, Roberto Mosca* and Patrick Aloy (*equal contributions).

JOURNAL   Chemistry & Biology.

TYPE   Review.

STAGE   Published.

CONTEXT   In the genome-scale metabolic models of cancer cell lines, and through the homology-based modeling of response to pH fluctuations, we showed how sequence-level information may be integrated into biological networks (Article 4). In this review, we brought this notion to the field of structure-based drug design where, through detailed atomistic analysis, researchers look for molecules that will exert systems-wide effects.

CITATION   Duran-Frigola et al. (2013) [226].

RESIDUES COEVOLUTION GUIDES THE SYSTEMATIC IDENTIFICATION OF ALTERNATIVE FUNCTIONAL CONFORMATIONS IN PROTEINS

ARTICLE 6   Dynamics and coevolution.

AUTHORS   Pedro Sfriso*, Miquel Duran-Frigola*, Roberto Mosca, Agustí Emperador, Patrick Aloy, Modesto Orozco (*equal contributions).

TYPE   Research article.

STAGE   Submitted.

CONTEXT   In Article 5 we highlighted the importance of large-scale structural analysis to enable systems pharmacology. To design drugs with desired molecular mechanisms of action, it is important to explore the conformational space of the target protein. In this work, we developed an automated protocol to discover functional conformers, which is one of the major concerns when looking for actionable regions in protein targets.

NOTE   This article contains supplementary material (Table S1) available upon request.

Sfriso P, Duran-Frigola M, Mosca R, Emperador A, Aloy P, Orozco M. Residues Coevolution Guides the Systematic Identification of Alternative Functional Conformations in Proteins. Structure. 2016 Jan 5;24(1):116-126. doi:10.1016/j.str.2015.10.025

DETECTING SIMILAR BINDING POCKETS TO ENABLE SYSTEMS PHARMACOLOGY

ARTICLE 7   Polypharmacology.

AUTHORS   Miquel Duran-Frigola*, Lydia Siragusa*, Gabriele Cruciani, Patrick Aloy (*equal contributions; author order to be decided).

TYPE   Research article.

STAGE   In preparation.

CONTEXT   Article 6 was aiming to expand the structural landscape of proteins. In this work, we analyzed the available structures to find putative binding pockets. Identifying similar binding sites in distinct proteins may enable polypharmacology, i.e. the simultaneous targeting of multiple proteins with a single drug. If achieved intendedly, the multi-target property may lead to drugs of broader impact on biological networks.

NOTE   This article is awaiting experimental validation.

Duran-Frigola M, Siragusa L, Ruppin E, Barril X, Cruciani G, Aloy P. Detecting similar binding pockets to enable systems polypharmacology. PLoS Comput Biol. 2017 Jun 29;13(6):e1005522. doi: 10.1371/journal.pcbi.1005522

Part III

DISCUSSION

GENERAL DISCUSSION

After all the high-throughput experiments conducted in the last years, perhaps the main lesson to be learnt is that human diseases have an extraordinarily complex molecular biology [1]. Together with this realization, it comes the admission that, in order to counteract disease complexity, sophisticated therapies are necessary, beyond the classical magic bullet idea [36, 256]. The switch towards more holistic therapies is insistently discussed in the literature, but the truth is that systems pharmacology is only in its infancy, and little has been achieved so far [226, 442]. To mature this field, it is key to bridge several levels of biological complexity, i.e. to gain translational power and merge the very detailed, abundant molecular data with the more sparse clinical outcomes.

The chemo-centric view of disease complements the more popular protein-centric view

A good source of clinical information are drug side effects, which can be viewed as direct phenotypic readouts of drug action (Article 8). Using side effects as *de facto* experimental measurements (Article 1), we performed a simple analysis to detect proteins, pathways, processes, and functions that are over-targeted among drugs causing the same adverse event. Doing this exercise for hundreds of side effects, we found a strong biological signal that could be used to gain mechanistic insights. In essence, our top-down study was analogous to routine omics experiments, where differential genetic factors are sought. Accordingly, while it was relatively easy to assign correlations between side effects and biological entities, these biological entities alone were poorly predictive of the clinical outcome. The emerging picture is that independent correlations will have to be combined with systemic knowledge in order to yield predictive models of biology.

As a complement to the mainstay molecular biology approach, we also mined chemical substructures that are over-represented in drugs causing a particular side effect. This approach requires a considerable amount of small molecule data, yet we found that, whenever applicable, it supplies remarkable predictive power. In a more thorough analysis of the chemotype-phenotype correlation, mainly expanded by environmental chemicals (Article 2), we confirmed that the ever-increasing volume of publicly available chemical data enables the chemo-centric view of health and disease. This view, while unaware of the underlying molecular mechanisms, can bridge gaps in biological knowledge (especially for predictive tasks) and uncovers relationships between diseases and treatments. In this regard, the global chemo-centric map of disease and health showed a striking coincidence with clinical data, and could accompany the more standard disease maps that are being produced based on shared molecular biology [154, 4].

Cell lines can be linked to biological networks to gain mechanistic insights and find novel therapeutic targets

In the era of large-scale experiments, the natural successors of the classical chemo-centric view are cell line screening panels. We did a retrospective analysis in the context of cancer therapy and observed that drugs approved for certain tumor types do not necessarily perform well in the corresponding cell lines (Article 3). Here again, it has been proposed that systems frameworks and molecular profiles are crucial to get the most of out of phenotypic screening [26]. Following this notion, we used genome-scale metabolic models as molecular descriptions of the cancer cell lines. In general, these models correlate well with phenotypic traits such as proliferation, and account for metabolic hallmarks [33]. Doing an effort to bring reductionist cognizance into the models, we developed a method to predict pH-activity profiles for every enzyme in the metabolic network (Article 4). The method benefitted from classical enzymology assays and sequence databases. Grouping the pH-activity profiles of all of the enzymes, we could reproduce the response of cancer cell lines to intracellular pH fluctuations. This exercise, which was continued to explore novel therapeutic strategies, exemplifies that, in the ground of systems biology, atomistic knowledge can be related to phenotypes.

Large-scale structural analysis enables systems-level perturbations on biological networks

More generally, we reviewed the recent attempts to embed molecular details into biological networks (Article 5). This is most relevant in the context of drug discovery, since the drug is an entity that needs to be designed with atomistic precision to exert, however, a systems-wide perturbation. In our opinion, this apparent paradox can be solved through automated, large-scale structural analysis. It is now possible to structurally annotate large biological networks [124], and it is in the agenda of drug design to profit more from the boom of structural data. In this regard, and to complement the conformational landscape of proteins, we developed an automated protocol to predict functional conformers solely based on the evolutionary history of the proteins and a structural snapshot (Article 6). In drug development, acknowledgement of functional conformers is important to apprehend the molecular mechanism of action and minimize the risk of attrition.

Particularly, it is key to mine putative binding sites across the conformational landscape. Identifying them in important conformations, possibly in allosteric regions, and in difficult areas like protein-protein interfaces is a timely topic in structure-based drug design (Article 5). Today, systematic detection of structural cavities is feasible at large, and similar binding sites have been found both between related and unrelated proteins. Binding site similarity explains the promiscuity of many drugs [126], which is often undesired due to adverse off-target reactions. Using systems biology tools, we have tried to turn this apparent disadvantage into a virtue —a subset of the promiscuous cases seem to

elicit compelling network perturbations (Article 7). Mining promising polypharmacology among the vast amount of possible multi-target combinations is a major requirement to materialize the move towards complex therapies.

Future outlook

During this latter work on polypharmacology (Article 7), where we integrated chemogenomics, sequence, structure, and network data, we understood that no blueprint exists to push atomistic information into biological networks, and that perturbations of these networks are difficult to translate to a phenotypic response. With the remarkable exception of genome-scale metabolic models [26], biological networks still lack a clear connection to phenotypes. It is feasible to have broad descriptions of disease complexity (Article 10 and Article 11), and differential data like gene expression can be incorporated into the networks, but to truly enable systems pharmacology solid quantitative methods are urged. Medicine is moving towards personalized treatments [443], molecular profiles are becoming cheaper, and new technologies don't cease to appear. More and more data will be produced, and systems pharmacology needs robust, validated protocols to undertake the challenge to digest them.

CONCLUSIONS

In this Thesis, we have performed several analyses at the intersection between chemoinformatics and bioinformatics, with a focus on systems biology and with the aim to address the translational problem of drug discovery. In a series of articles, we have reached the following conclusions:

- Using drug side effects as phenotypic readouts of drug action, many correlations between drug (off)targets and the eventual clinical outcome can be learnt. However, this mechanistic information has limited predictive power (Article 1).

- Chemical information, on the other hand, is much more difficult to extract, but it provides considerable predictive power (Article 1 and Article 2).

- In general, a chemo-centric view of health and disease is today feasible with the available small molecule data. This view uncovers clinically-relevant connections between diseases and therapies (Article 2).

- The modern realization of the chemo-centric approach to pharmacology are cell-based assays. We observed that response in cancer cell lines is not tissue-specific, suggesting that these panels need to be characterized with biological profiles (Article 3).

- By using genome-scale metabolic models to pursue this characterization, we could predict cell-line response to intracellular pH fluctuations. The dependence on pH was incorporated into the metabolic models using reductionist enzymology information and sequence data (Article 4).

- The results in the pH-sensitive metabolic models show that systems biology frameworks can be informed with detailed molecular data. This is particularly important in drug discovery, where structural information on protein targets is crucial to enable drug design (Article 5).

- To contribute to a richer characterization of the structural landscape, we developed a protocol to predict alternative functional conformers of proteins. This method greatly benefitted from sequence data to learn evolutionary histories (Article 6).

- Within the available structural landscape, thousands of putative binding cavities could be identified. Similar binding sites were found between both related and unrelated proteins, enabling the polypharmacology idea by means of network-based simulations (Article 7).

Part IV

APPENDIX

A

SUPPLEMENTARY ARTICLES

A.1 RECYCLING SIDE EFFECTS INTO CLINICAL MARKERS FOR DRUG REPOSITIONING

ARTICLE 8   Recycling side effects.

AUTHORS   Miquel Duran-Frigola and Patrick Aloy.

JOURNAL   Genome Medicine.

TYPE   Short comment.

STAGE   Published.

CITATION   Duran-Frigola et al. (2012) [160].

NOTE   This article is a short perspective comment, available at `http://dx.doi.org/10.1186/gm302`.

ABSTRACT

Side-effects are the unintended consequence of therapeutic treatments, but they can also be seen as valuable read-outs of drug effects in humans; these effects are difficult to infer or predict from pre-clinical models. Indeed, some studies suggest that drugs with similar side-effect profiles may also share therapeutic properties through related mechanisms of action. A recent publication exploits this concept to systematically investigate new indications for already marketed drugs, and presents a strategy to get the most out of the tiny portion of chemicals that have proved to be effective and safe.

A.2 INTSIDE: A WEB SERVER FOR THE CHEMICAL AND BIOLOGICAL EXAMINATION OF DRUG SIDE EFFECTS

ARTICLE 9   IntSide webserver.

AUTHORS   Teresa Juan-Blanco, Miquel Duran-Frigola and Patrick Aloy.

JOURNAL   Bioinformatics.

TYPE   Research article.

STAGE   Published.

CITATION   Juan-Blanco et al. (2015) [444].

AVAILABLILITY   IntSide is available at `intside.irbbarcelona.org`.

NOTE   This is a resource based on Article 1. The article is available at `http://dx.doi.org/10.1093/bioinformatics/btu688`.

ABSTRACT

Drug side effects are one of the main health threats worldwide, and an important obstacle in drug development. Understanding how adverse reactions occur requires knowledge on drug mechanisms at the molecular level. Despite recent advances, the need for tools and methods that facilitate side effect anticipation still remains. Here, we present IntSide, a web server that integrates chemical and biological information to elucidate the molecular mechanisms underlying drug side effects. IntSide currently catalogs $1,175$ side effects caused by 996 drugs, associated with drug features divided into eight categories, belonging to either biology or chemistry. On the biological side, IntSide reports drug targets and off-targets, pathways, molecular functions and biological processes. From a chemical viewpoint, it includes molecular fingerprints, scaffolds and chemical entities. Finally, we also integrate additional biological data, such as protein interactions and disease-related genes, to facilitate mechanistic interpretations.

A.3   CHARTING THE MOLECULAR LINKS BETWEEN DRIVER AND SUSCEPTIBILITY GENES IN COLORECTAL CANCER

ARTICLE 10   Colorectal cancer intractome.

AUTHORS   Rodrigo Arroyo, Miquel Duran-Frigola, Clara Berenguer, Montse Soler-López and Patrick Aloy.

JOURNAL   Biochemical and Biophysical Research Communications.

TYPE   Research article.

STAGE   Published.

CITATION   Arroyo et al. (2014) [445].

NOTE   The article and the supplementary information are available at `http://dx.doi.org/10.1016/j.bbrc.2013.12.012`.

ABSTRACT

Despite significant advances in the identification of specific genes and pathways important in the onset and progression of colorectal cancer (CRC), mechanistic insight into the relationship between driver and susceptibility genes is needed. In this paper, we systematically explore physical interactions between causative and putative CRC susceptibility genes to reveal the molecular mechanisms involved in tumor biology. In total, we identify 622 high-confidence protein-protein interactions between 42 CRC causative and 65 candidate susceptibility genes. Among the latter, 28 are located in the CRCS9 loci, related to the etiology of CRC, and 17 are co-expressed with well-established CRC drivers, which makes them excellent candidates for further functional studies. Moreover, we find a high degree of functional coherence between connected driver and susceptibility genes, which indicates that our network-based strategy is useful to gain insight into the underlying mechanisms of those proteins with unknown roles in CRC.

A.4 SYSTEMATIC IDENTIFICATION OF MOLECULAR LINKS BETWEEN CORE AND CANDIDATE GENES IN BREAST CANCER

ARTICLE 11   Breast cancer interactome.

AUTHORS   Rodrigo Arroyo, Guillermo Suñé, Andreas Zanzoni, Miquel Duran-Frigola, Víctor Alcalde, Travis H. Stracker, Montse Soler-López and Patrick Aloy.

ABSTRACT

Despite the remarkable progress achieved in the identification of specific genes involved in breast cancer (BC), our understanding of their complex functioning is still limited. In this manuscript, we systematically explore the existence of direct physical interactions between the products of BC core and associated genes. Our aim is to generate a protein interaction network of BC-associated gene products and suggest potential molecular mechanisms to unveil their role

in the disease. In total, we report 599 novel high-confidence interactions among 44 BC core, 54 BC candidate/associated and 96 newly identified proteins. Our findings indicate that this network-based approach is indeed a robust inference tool to pinpoint new potential players and gain insight into the underlying mechanisms of those proteins with previously unknown roles in BC. To illustrate the power of our approach, we provide initial validation of two BC-associated proteins on the alteration of DNA damage response as a result of specific re-wiring interactions. Overall, our BC-related network may serve as a framework to integrate clinical and molecular data and foster novel global therapeutic strategies.

A.5   ISOLATION OF HUMAN COLON STEM CELLS USING SURFACE EXPRESSION OF PROTEIN TYROSINE KINASE RECEPTOR 7 (PTK7)

ARTICLE 12   Isolation of human colon stem cells.

AUTHORS   Peter Jung, Christian Sommer, Francisco Barriga, Miquel Duran-Frigola, Patrick Aloy, Matthias Selbach, Doug Winton and Eduard Batlle.

STAGE   Submitted.

OUR CONTRIBUTION   We did an enrichment analysis of membrane proteins. This article is outside the focus of the Thesis.

ABSTRACT

Long-term growth and multi-lineage differentiation in the intestinal tract is sustained by a population of actively dividing stem cells. Here we identify Protein Kinase 7 (PTK7) as a highly specific marker for human colonic stem cells (hCoSCs). PTK7 surface abundance readily enables purification of hCoSCs from organoid cultures and patient-derived mucosa samples. PTK7+ hCoSCs display combined features of the Lgr5+ Intestinal Stem Cells and crypt Label Retaining Cells (LRCs).

A.6   SYSTEMS-WIDE PREDICTION OF GENE PROMISCUITY REVEALS A NEW UNDERGROUND ALTERNATIVE ROUTE FOR PYRIDOXAL 5′-PHOSPHATE PRODUCTION IN E. COLI

ARTICLE 13   Systems-wide prediction of gene promiscuity.

AUTHORS   Matthew A. Oberhardt, Raphy Zarecki, Leah Reshef, Miquel Duran-Frigola, Rachel Schreiber, Cristopher S. Henry, Nir Ben-Tal, Daniel J. Dwyer, Uri Gophna, Eytan Ruppin.

STAGE   Submitted.

OUR CONTRIBUTION   We did an structural analysis to characterize a promiscuous enzymatic function. This article is outside the focus of the Thesis.

ABSTRACT

Recent insights suggest that non-specific and/or promiscuous enzymes are common and active across life. Understanding the role of such enzymes is an important open question in biology. Here we develop a genome-wide method, PROPER, that uses a permissive PSI-BLAST approach to predict promiscuous activities of metabolic genes. Gene promiscuity is typically studied experimentally using multicopy suppression, in which over-expression of a promiscuous 'replacer' gene rescues lethality caused by inactivation of a 'target' gene. We use PROPER to predict multicopy suppression in *Escherichia coli*, achieving highly significant overlap with published cases (hypergeometric $P = 4.4 \times 10^{-13}$). We then validate three novel predicted target-replacer gene pairs in new multicopy expression experiments. An analysis of overexpression lethality and co-evolution trends in predicted target-replacer pairs suggests that promiscuity plays a functional backup role in cells. We next go beyond PROPER and develop a network-based approach, GEM-PROPER, that integrates PROPER with genome-scale metabolic modeling to predict promiscuous replacements via alternative metabolic pathways. GEM-PROPER predicts a new indirect replacer (*thiG*) for an essential enzyme (*pdxB*) in production of pyridoxal 5'-phosphate (the active form of Vitamin B6), which we validate experimentally via multicopy suppression. We perform a structural analysis of *thiG* to determine its potential promiscuous active site, which we validate experimentally by inactivating the pertaining residues and showing a loss of replacer activity. Thus, this study is a successful example where a computational investigation leads to a network-based identification of an indirect promiscuous replacement of a key metabolic enzyme, which would have been extremely difficult to identify directly.

[1] A.-L. Barabási, N. Gulbahce, and J. Loscalzo, "Network medicine: a network-based approach to human disease.," *Nature reviews. Genetics*, vol. 12, pp. 56–68, Jan. 2011.

[2] L. I. Furlong, "Human diseases through the lens of network biology.," *Trends in genetics : TIG*, vol. 29, pp. 150–9, Mar. 2013.

[3] A. Zanzoni, M. Soler-López, and P. Aloy, "A network medicine approach to human disease," *FEBS Letters*, vol. 583, pp. 1759–1765, June 2009.

[4] J. Menche, A. Sharma, M. Kitsak, S. D. Ghiassian, M. Vidal, J. Loscalzo, and A.-L. Barabási, "Disease networks. Uncovering disease-disease relationships through the incomplete interactome.," *Science (New York, N.Y.)*, vol. 347, p. 1257601, Feb. 2015.

[5] T. Ideker and R. Sharan, "Protein networks in disease.," *Genome research*, vol. 18, pp. 644–52, Apr. 2008.

[6] M. G. Kann, "Protein interactions and disease: computational approaches to uncover the etiology of diseases.," *Briefings in bioinformatics*, vol. 8, pp. 333–46, Sept. 2007.

[7] E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, and e. a. Zody, "Initial sequencing and analysis of the human genome.," *Nature*, vol. 409, pp. 860–921, Feb. 2001.

[8] J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, and e. a. Mural, "The sequence of the human genome.," *Science (New York, N.Y.)*, vol. 291, pp. 1304–51, Feb. 2001.

[9] F. Mazzocchi, "Complexity in biology. Exceeding the limits of reductionism and determinism using complexity theory.," *EMBO reports*, vol. 9, pp. 10–4, Jan. 2008.

[10] T. Potthast, "Paradigm shifts versus fashion shifts? Systems and synthetic biology as new epistemic entities in understanding and making 'life'.," *EMBO reports*, vol. 10 Suppl 1, pp. S42–5, Aug. 2009.

[11] H. V. Westerhoff and B. O. Palsson, "The evolution of molecular biology into systems biology.," *Nature biotechnology*, vol. 22, pp. 1249–52, Oct. 2004.

[12] L. Onsager, "Reciprocal Relations in Irreversible Processes. I.," *Physical Review*, vol. 37, pp. 405–426, Feb. 1931.

[13] L. Onsager, "Reciprocal Relations in Irreversible Processes. II.," *Physical Review*, vol. 38, pp. 2265–2279, Dec. 1931.

[14] H. Rottenberg, S. R. Caplan, and A. Essig, "Soichiometry and coupling: theories of oxidative phosphorylation.," *Nature*, vol. 216, pp. 610–1, Nov. 1967.

[15] T. Rolland, M. Tasan, B. Charloteaux, S. J. Pevzner, and e. a. Zhong, "A Proteome-Scale Map of the Human Interactome Network," *Cell*, vol. 159, pp. 1212–1226, Nov. 2014.

[16] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, pp. 47–97, Jan. 2002.

[17] X. Zhu, M. Gerstein, and M. Snyder, "Getting connected: analysis and principles of biological networks.," *Genes & development*, vol. 21, pp. 1010–24, May 2007.

[18] G. Caldarelli, *Scale-Free Networks: Complex Webs in Nature and Technology*. Oxford University Press, 2007.

[19] S. Wachi, K. Yoneda, and R. Wu, "Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues.," *Bioinformatics (Oxford, England)*, vol. 21, pp. 4205–8, Dec. 2005.

[20] J. Xu and Y. Li, "Discovering disease-genes by topological features in human protein-protein interaction network.," *Bioinformatics (Oxford, England)*, vol. 22, pp. 2800–5, Nov. 2006.

[21] K. Venkatesan, J.-F. Rual, A. Vazquez, U. Stelzl, I. Lemmens, T. Hirozane-Kishikawa, T. Hao, M. Zenkner, X. Xin, K.-I. Goh, M. A. Yildirim, N. Simonis, K. Heinzmann, F. Gebreab, J. M. Sahalie, S. Cevik, C. Simon, A.-S. de Smet, E. Dann, A. Smolyar, A. Vinayagam, H. Yu, D. Szeto, H. Borick, A. Dricot, N. Klitgord, R. R. Murray, C. Lin, M. Lalowski, J. Timm, K. Rau, C. Boone, P. Braun, M. E. Cusick, F. P. Roth, D. E. Hill, J. Tavernier, E. E. Wanker, A.-L. Barabási, and M. Vidal, "An empirical framework for binary interactome mapping.," *Nature methods*, vol. 6, pp. 83–90, Jan. 2009.

[22] J. M. Whitacre, "Biological robustness: paradigms, mechanisms, and systems principles.," *Frontiers in genetics*, vol. 3, p. 67, Jan. 2012.

[23] H. Kitano, "Biological robustness.," *Nature reviews. Genetics*, vol. 5, pp. 826–37, Nov. 2004.

[24] H. Kitano, "Cancer as a robust system: implications for anticancer therapy.," *Nature reviews. Cancer*, vol. 4, pp. 227–35, Mar. 2004.

[25] H. Kitano, "Towards a theory of biological robustness.," *Molecular systems biology*, vol. 3, p. 137, Jan. 2007.

[26] A. Bordbar, J. M. Monk, Z. A. King, and B. O. Palsson, "Constraint-based models predict metabolic and associated cellular functions.," *Nature reviews. Genetics*, vol. 15, pp. 107–20, Feb. 2014.

[27] A. M. Feist, M. J. Herrgå rd, I. Thiele, J. L. Reed, and e. a. Palsson, "Reconstruction of biochemical networks in microorganisms.," *Nature reviews. Microbiology*, vol. 7, pp. 129–43, Feb. 2009.

[28] I. Thiele and B. O. Palsson, "A protocol for generating a high-quality genome-scale metabolic reconstruction.," *Nature protocols*, vol. 5, pp. 93–121, Jan. 2010.

[29] N. C. Duarte, S. A. Becker, N. Jamshidi, I. Thiele, and e. a. Mo, "Global reconstruction of the human metabolic network based on genomic and bibliomic data.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, pp. 1777–82, Feb. 2007.

[30] I. Thiele, N. Swainston, R. M. T. Fleming, A. Hoppe, and e. a. Sahoo, "A community-driven global reconstruction of human metabolism.," *Nature biotechnology*, vol. 31, pp. 419–25, May 2013.

[31] J. D. Orth, I. Thiele, and B. O. Palsson, "What is flux balance analysis?," *Nature biotechnology*, vol. 28, pp. 245–8, Mar. 2010.

[32] J. S. Edwards, R. U. Ibarra, and B. O. Palsson, "In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data.," *Nature biotechnology*, vol. 19, pp. 125–30, Feb. 2001.

[33] K. Yizhak, E. Gaude, S. Le Dévédec, Y. Y. Waldman, G. Y. Stein, B. van de Water, C. Frezza, and E. Ruppin, "Phenotype-based cell-specific metabolic modeling reveals metabolic liabilities of cancer.," *eLife*, vol. 3, Jan. 2014.

[34] J. Drews, "Case histories, magic bullets and the state of drug discovery.," *Nature reviews. Drug discovery*, vol. 5, pp. 635–40, Aug. 2006.

[35] V. A. Narayan, M. Mohwinckel, G. Pisano, M. Yang, and e. a. Manji, "Beyond magic bullets: true innovation in health care.," *Nature reviews. Drug discovery*, vol. 12, pp. 85–6, Feb. 2013.

[36] F. Pammolli, L. Magazzini, and M. Riccaboni, "The productivity crisis in pharmaceutical R&D.," *Nature reviews. Drug discovery*, vol. 10, pp. 428–38, June 2011.

[37] D. C. Swinney and J. Anthony, "How were new medicines discovered?," *Nature reviews. Drug discovery*, vol. 10, pp. 507–19, July 2011.

[38] M. A. Lindsay, "Target discovery.," *Nature reviews. Drug discovery*, vol. 2, pp. 831–8, Oct. 2003.

[39] P. Imming, C. Sinning, and A. Meyer, "Drugs, their targets and the nature and number of drug targets.," *Nature reviews. Drug discovery*, vol. 5, pp. 821–34, Oct. 2006.

[40] J. P. Overington, B. Al-Lazikani, and A. L. Hopkins, "How many drug targets are there?," *Nature reviews. Drug discovery*, vol. 5, pp. 993–6, Dec. 2006.

[41] T. Fulmer, "Rethinking mechanisms of drug discovery," *Science-Business eXchange*, vol. 4, July 2011.

[42] M. P. Gleeson, A. Hersey, D. Montanari, and J. Overington, "Probing the links between in vitro potency, ADMET and physicochemical parameters.," *Nature reviews. Drug discovery*, vol. 10, pp. 197–208, Mar. 2011.

[43] M. Congreve, C. W. Murray, and T. L. Blundell, "Structural biology and drug discovery.," *Drug discovery today*, vol. 10, pp. 895–907, July 2005.

[44] C. R. Beddell, P. J. Goodford, F. E. Norrington, S. Wilkinson, and e. a. Wootton, "Compounds designed to fit a site of known structure in human haemoglobin.," *British journal of pharmacology*, vol. 57, pp. 201–9, June 1976.

[45] P. J. Goodford, J. St-Louis, and R. Wootton, "The interaction of human haemoglobin with allosteric effectors as a model for drug-receptor interactions.," *British journal of pharmacology*, vol. 68, pp. 741–8, Apr. 1980.

[46] T. L. Blundell, B. L. Sibanda, M. J. Sternberg, and J. M. Thornton, "Knowledge-based prediction of protein structures and the design of novel molecules.," *Nature*, vol. 326, pp. 347–52, Jan. 1987.

[47] U. Heinemann, K. Büssow, U. Mueller, and P. Umbach, "Facilities and methods for the high-throughput crystal structural analysis of human proteins.," *Accounts of chemical research*, vol. 36, pp. 157–63, Mar. 2003.

[48] S. A. Lesley, P. Kuhn, A. Godzik, A. M. Deacon, and e. a. Mathews, "Structural genomics of the Thermotoga maritima proteome implemented in a high-throughput structure determination pipeline.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 11664–9, Sept. 2002.

[49] R. F. Service, "Structural genomics. Tapping DNA for structures produces a trickle.," *Science (New York, N.Y.)*, vol. 298, pp. 948–50, Nov. 2002.

[50] J. M. Thompson, N. G. Sgourakis, G. Liu, P. Rossi, and e. a. Tang, "Accurate protein structure modeling using sparse NMR data and homologous structure information.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, pp. 9875–80, June 2012.

[51] E. Krieger, K. Joo, J. Lee, J. Lee, and e. a. Raman, "Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8.," *Proteins*, vol. 77 Suppl 9, pp. 114–22, Jan. 2009.

[52] M. Gao and J. Skolnick, "A comprehensive survey of small-molecule binding pockets in proteins.," *PLoS computational biology*, vol. 9, p. e1003302, Oct. 2013.

[53] R. D. Taylor, P. J. Jewsbury, and J. W. Essex, "A review of protein-small molecule docking methods.," *Journal of computer-aided molecular design*, vol. 16, pp. 151–66, Mar. 2002.

[54] H. M. Geysen, F. Schoenen, D. Wagner, and R. Wagner, "Combinatorial compound libraries for drug discovery: an ongoing challenge.," *Nature reviews. Drug discovery*, vol. 2, pp. 222–30, Mar. 2003.

[55] S. Surade and T. L. Blundell, "Structural biology and drug discovery of difficult targets: the limits of ligandability.," *Chemistry & biology*, vol. 19, pp. 42–50, Jan. 2012.

[56] A. L. Hopkins and C. R. Groom, "The druggable genome.," *Nature reviews. Drug discovery*, vol. 1, pp. 727–30, Sept. 2002.

[57] F. N. B. Edfeldt, R. H. A. Folmer, and A. L. Breeze, "Fragment screening to predict druggability (ligandability) and lead discovery success.," *Drug discovery today*, vol. 16, pp. 284–7, Apr. 2011.

[58] M. J. Keiser, J. J. Irwin, and B. K. Shoichet, "The chemical basis of pharmacology.," *Biochemistry*, vol. 49, pp. 10267–76, Dec. 2010.

[59] M. J. Keiser, B. L. Roth, B. N. Armbruster, P. Ernsberger, and e. a. Irwin, "Relating protein pharmacology by ligand chemistry.," *Nature biotechnology*, vol. 25, pp. 197–206, Feb. 2007.

[60] M. J. Keiser, V. Setola, J. J. Irwin, C. Laggner, and e. a. Abbas, "Predicting new molecular targets for known drugs.," *Nature*, vol. 462, pp. 175–81, Nov. 2009.

[61] H. Lin, M. F. Sassano, B. L. Roth, and B. K. Shoichet, "A pharmacological organization of G protein-coupled receptors.," *Nature methods*, vol. 10, pp. 140–6, Feb. 2013.

[62] E. Lounkine, M. J. Keiser, S. Whitebread, D. Mikhailov, and e. a. Hamon, "Large-scale prediction and testing of drug activity on side-effect targets.," *Nature*, vol. 486, pp. 361–7, June 2012.

[63] E. Gregori-Puigjané and J. Mestres, "Coverage and bias in chemical library design.," *Current opinion in chemical biology*, vol. 12, pp. 359–65, June 2008.

[64] A. P. Davis, C. G. Murphy, R. Johnson, J. M. Lay, and e. a. Lennon-Hopkins, "The Comparative Toxicogenomics Database: update 2013.," *Nucleic acids research*, vol. 41, pp. D1104–14, Jan. 2013.

[65] L. J. Bellis, R. Akhtar, B. Al-Lazikani, F. Atkinson, and e. a. Bento, "Collation and data-mining of literature bioactivity data for drug discovery.," *Biochemical Society transactions*, vol. 39, pp. 1365–70, Oct. 2011.

[66] K. Strebhardt and A. Ullrich, "Paul Ehrlich's magic bullet concept: 100 years of progress.," *Nature reviews. Cancer*, vol. 8, pp. 473–80, June 2008.

[67] F. Bosch and L. Rosich, "The contributions of Paul Ehrlich to pharmacology: a tribute on the occasion of the centenary of his Nobel Prize.," *Pharmacology*, vol. 82, pp. 171–9, Jan. 2008.

[68] Y. Lazebnik, "Can a biologist fix a radio?–Or, what I learned while studying apoptosis.," *Cancer cell*, vol. 2, pp. 179–82, Sept. 2002.

[69] M. Duran-Frigola and P. Aloy, "Analysis of chemical and biological features yields mechanistic insights into drug side effects.," *Chemistry & biology*, vol. 20, pp. 594–603, Apr. 2013.

[70] K. M. Giacomini, R. M. Krauss, D. M. Roden, M. Eichelbaum, and e. a. Hayden, "When good drugs go bad.," *Nature*, vol. 446, pp. 975–7, Apr. 2007.

[71] T.-Y. Wu, M.-H. Jen, A. Bottle, M. Molokhia, and e. a. Aylin, "Ten-year trends in hospital admissions for adverse drug reactions in England 1999-2009.," *Journal of the Royal Society of Medicine*, vol. 103, pp. 239–50, June 2010.

[72] N. P. Tatonetti, P. P. Ye, R. Daneshjou, and R. B. Altman, "Data-driven prediction of drug effects and interactions.," *Science translational medicine*, vol. 4, p. 125ra31, Mar. 2012.

[73] M. Kuhn, M. Campillos, I. Letunic, L. J. Jensen, and e. a. Bork, "A side effect resource to capture phenotypic effects of drugs.," *Molecular systems biology*, vol. 6, p. 343, Jan. 2010.

[74] M. Campillos, M. Kuhn, A.-C. Gavin, L. J. Jensen, and e. a. Bork, "Drug target identification using side-effect similarity.," *Science (New York, N.Y.)*, vol. 321, pp. 263–6, July 2008.

[75] I. Wallach, N. Jaitly, and R. Lilien, "A structure-based approach for mapping adverse drug reactions to the perturbation of underlying biological pathways.," *PloS one*, vol. 5, p. e12063, Jan. 2010.

[76] S. Lee, K. H. Lee, M. Song, and D. Lee, "Building the process-drug-side effect network to discover the relationship between biological processes and side effects.," *BMC bioinformatics*, vol. 12 Suppl 2, p. S2, Jan. 2011.

[77] P. Holzer, "Opioid receptors in the gastrointestinal tract.," *Regulatory peptides*, vol. 155, pp. 11–7, June 2009.

[78] A. Vedani, M. Dobler, and M. Smieško, "VirtualToxLab - a platform for estimating the toxic potential of drugs, chemicals and natural products.," *Toxicology and applied pharmacology*, vol. 261, pp. 142–53, June 2012.

[79] A. Bauer-Mehren, E. M. van Mullingen, P. Avillach, M. D. C. Carrascosa, R. Garcia-Serna, J. Piñero, B. Singh, P. Lopes, J. L. Oliveira, G. Diallo, E. Ahlberg Helgee, S. Boyer, J. Mestres, F. Sanz, J. A. Kors, and L. I. Furlong, "Automatic filtering and substantiation of drug safety signals.," *PLoS computational biology*, vol. 8, p. e1002457, Jan. 2012.

[80] R. Garcia-Serna, O. Ursu, T. I. Oprea, and J. Mestres, "iPHACE: integrative navigation in pharmacological space.," *Bioinformatics (Oxford, England)*, vol. 26, pp. 985–6, Apr. 2010.

[81] J. Mestres, L. Martín-Couce, E. Gregori-Puigjané, M. Cases, and S. Boyer, "Ligand-based approach to in silico pharmacology: nuclear receptor profiling.," *Journal of chemical information and modeling*, vol. 46, pp. 2725–36, Jan. 2006.

[82] M. E. Hillenmeyer, E. Ericson, R. W. Davis, C. Nislow, and e. a. Koller, "Systematic analysis of genome-wide fitness data in yeast reveals novel gene function and drug action.," *Genome biology*, vol. 11, p. R30, Jan. 2010.

[83] D. P. Williams and B. K. Park, "Idiosyncratic toxicity: the role of toxicophores and bioactivation.," *Drug discovery today*, vol. 8, pp. 1044–50, Nov. 2003.

[84] A. Bender, J. Scheiber, M. Glick, J. W. Davies, and e. a. Azzaoui, "Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure.," *ChemMedChem*, vol. 2, pp. 861–73, June 2007.

[85] J. Scheiber, J. L. Jenkins, S. C. K. Sukuru, A. Bender, and e. a. Mikhailov, "Mapping adverse drug reactions in chemical space.," *Journal of medicinal chemistry*, vol. 52, pp. 3103–7, May 2009.

[86] A. Cami, A. Arnold, S. Manzi, and B. Reis, "Predicting adverse drug events using pharmacological network models.," *Science translational medicine*, vol. 3, p. 114ra127, Dec. 2011.

[87] A. Gottlieb, G. Y. Stein, E. Ruppin, and R. Sharan, "PREDICT: a method for inferring novel drug indications with application to personalized medicine.," *Molecular systems biology*, vol. 7, p. 496, Jan. 2011.

[88] E. Pauwels, V. Stoven, and Y. Yamanishi, "Predicting drug side-effect profiles: a chemical fragment-based approach.," *BMC bioinformatics*, vol. 12, p. 169, Jan. 2011.

[89] T. I. Oprea and J. Gottfries, "Chemography: the art of navigating in chemical space.," *Journal of combinatorial chemistry*, vol. 3, pp. 157–66, Jan. 2001.

[90] T. Varin, A. Schuffenhauer, P. Ertl, and S. Renner, "Mining for bioactive scaffolds with scaffold networks: improved compound set enrichment from primary screening data.," *Journal of chemical information and modeling*, vol. 51, pp. 1528–38, July 2011.

[91] G. W. Bemis and M. A. Murcko, "The properties of known drugs. 1. Molecular frameworks.," *Journal of medicinal chemistry*, vol. 39, pp. 2887–93, July 1996.

[92] S. Wetzel, K. Klein, S. Renner, D. Rauh, and e. a. Oprea, "Interactive exploration of chemical space with Scaffold Hunter.," *Nature chemical biology*, vol. 5, pp. 581–3, Aug. 2009.

[93] P. de Matos, R. Alcántara, A. Dekker, M. Ennis, J. Hastings, K. Haug, I. Spiteri, S. Turner, and C. Steinbeck, "Chemical Entities of Biological Interest: an update.," *Nucleic acids research*, vol. 38, pp. D249–54, Jan. 2010.

[94] J. Mestres, E. Gregori-Puigjané, S. Valverde, and R. V. Solé, "Data completeness–the Achilles heel of drug-target networks.," *Nature biotechnology*, vol. 26, pp. 983–4, Sept. 2008.

[95] A. Trifilieff, A. Da Silva, and J. P. Gies, "Kinins and respiratory tract diseases.," *The European respiratory journal*, vol. 6, pp. 576–87, Apr. 1993.

[96] L. M. C. Vasquez-Pinto, F. Nantel, P. Sirois, and S. Jancar, "Bradykinin B(1) receptor antagonist R954 inhibits eosinophil activation/proliferation/migration and increases TGF-beta and VEGF in a murine model of asthma.," *Neuropeptides*, vol. 44, pp. 107–13, Apr. 2010.

[97] N. Senanayake and G. C. Román, "Disorders of neuromuscular transmission due to natural environmental toxins.," *Journal of the neurological sciences*, vol. 107, pp. 1–13, Jan. 1992.

[98] J. Clayton-Smith, S. Walters, E. Hobson, E. Burkitt-Wright, and e. a. Smith, "Xq28 duplication presenting with intestinal and bladder dysfunction and a distinctive facial appearance.," *European journal of human genetics : EJHG*, vol. 17, pp. 434–43, Apr. 2009.

[99] R. Lin, K. Karpa, N. Kabbani, P. Goldman-Rakic, and e. a. Levenson, "Dopamine D2 and D3 receptors are linked to the actin cytoskeleton via interaction with filamin A.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, pp. 5258–63, Apr. 2001.

[100] G. A. McQuibban, J.-H. Gong, J. P. Wong, J. L. Wallace, and e. a. Clark-Lewis, "Matrix metalloproteinase processing of monocyte chemoattractant proteins generates CC chemokine receptor antagonists with anti-inflammatory properties in vivo.," *Blood*, vol. 100, pp. 1160–7, Aug. 2002.

[101] R. Vallon, R. Müller, D. Moosmayer, E. Gerlach, and P. Angel, "The catalytic domain of activated collagenase I (MMP-1) is absolutely required for interaction with its specific inhibitor, tissue inhibitor of metalloproteinases-1 (TIMP-1).," *European journal of biochemistry / FEBS*, vol. 244, pp. 81–8, Feb. 1997.

[102] M. Shuvy, S. Abedat, R. Beeri, M. Valitsky, and e. a. Daher, "Raloxifene attenuates Gas6 and apoptosis in experimental aortic valve disease in renal failure.," *American journal of physiology. Heart and circulatory physiology*, vol. 300, pp. H1829–40, May 2011.

[103] A. Fura, "Role of pharmacologically active metabolites in drug discovery and development.," *Drug discovery today*, vol. 11, pp. 133–42, Feb. 2006.

[104] M. Soler-López, A. Zanzoni, R. Lluís, U. Stelzl, and P. Aloy, "Interactome mapping suggests new mechanistic details underlying Alzheimer's disease.," *Genome research*, vol. 21, pp. 364–76, Mar. 2011.

[105] W. N. Arnold and L. S. Loftus, "Xanthopsia and van Gogh's yellow palette.," *Eye (London, England)*, vol. 5 (Pt 5), pp. 503–10, Jan. 1991.

[106] B. M. Tune and D. Fravert, "Mechanisms of cephalosporin nephrotoxicity: a comparison of cephaloridine and cephaloglycin.," *Kidney international*, vol. 18, pp. 591–600, Nov. 1980.

[107] P. J. Blanchet, M.-T. Parent, P. H. Rompré, and D. Lévesque, "Relevance of animal models to human tardive dyskinesia.," *Behavioral and brain functions : BBF*, vol. 8, p. 12, Jan. 2012.

[108] S. I. Berger and R. Iyengar, "Role of systems pharmacology in understanding drug adverse events.," *Wiley interdisciplinary reviews. Systems biology and medicine*, vol. 3, pp. 129–35, Jan. 2011.

[109] L. Becquemont, "Pharmacogenomics of adverse drug reactions: practical applications and perspectives.," *Pharmacogenomics*, vol. 10, pp. 961–9, June 2009.

[110] D. Gurwitz and A. G. Motulsky, "'Drug reactions, enzymes, and biochemical genetics': 50 years later.," *Pharmacogenomics*, vol. 8, pp. 1479–84, Nov. 2007.

[111] C. Knox, V. Law, T. Jewison, P. Liu, and e. a. Ly, "DrugBank 3.0: a comprehensive resource for 'omics' research on drugs.," *Nucleic acids research*, vol. 39, pp. D1035–41, Jan. 2011.

[112] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, and e. a. Vandermeersch, "Open Babel: An open chemical toolbox.," *Journal of cheminformatics*, vol. 3, p. 33, Jan. 2011.

[113] M. Kuhn, D. Szklarczyk, A. Franceschini, C. von Mering, and e. a. Jensen, "STITCH 3: zooming in on protein-chemical interactions.," *Nucleic acids research*, vol. 40, pp. D876–80, Jan. 2012.

[114] M. Kanehisa and S. Goto, "KEGG: kyoto encyclopedia of genes and genomes.," *Nucleic acids research*, vol. 28, pp. 27–30, Jan. 2000.

[115] R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, and e. a. Dettling, "Bioconductor: open software development for computational biology and bioinformatics.," *Genome biology*, vol. 5, p. R80, Jan. 2004.

[116] J. a. D. Ferreira and F. M. Couto, "Semantic similarity for automatic classification of chemical compounds.," *PLoS computational biology*, vol. 6, Jan. 2010.

[117] A. Alexa, J. Rahnenfuhrer, and T. Lengauer, "Improved scoring of functional groups from gene expression data by decorrelating GO graph structure," *Bioinformatics*, vol. 22, pp. 1600–1607, Apr. 2006.

[118] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *The Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, Feb. 2011.

[119] C. Pesquita, D. Faria, A. O. Falcão, P. Lord, and F. M. Couto, "Semantic similarity in biomedical ontologies.," *PLoS computational biology*, vol. 5, p. e1000443, July 2009.

[120] W. W. M. Fleuren, S. Verhoeven, R. Frijters, B. Heupers, and e. a. Polman, "CoPub update: CoPub 5.0 a text mining system to answer biological questions.," *Nucleic acids research*, vol. 39, pp. W450–4, July 2011.

[121] N. F. Noy, N. H. Shah, P. L. Whetzel, B. Dai, and e. a. Dorf, "BioPortal: ontologies and integrated data resources at the click of a mouse.," *Nucleic acids research*, vol. 37, pp. W170–3, July 2009.

[122] R. Frijters, M. van Vugt, R. Smeets, R. van Schaik, J. de Vlieg, and W. Alkema, "Literature mining for the discovery of hidden connections between drugs, genes and diseases.," *PLoS computational biology*, vol. 6, Jan. 2010.

[123] A. Bauer-Mehren, M. Rautschka, F. Sanz, and L. I. Furlong, "DisGeNET: a Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks.," *Bioinformatics (Oxford, England)*, vol. 26, pp. 2924–6, Nov. 2010.

[124] R. Mosca, A. Céol, and P. Aloy, "Interactome3D: adding structural details to protein networks.," *Nature methods*, vol. 10, pp. 47–53, Jan. 2013.

[125] M. Duran-Frigola, D. Rossell, and P. Aloy, "A chemo-centric view of human health and disease.," *Nature communications*, vol. 5, p. 5676, Jan. 2014.

[126] M. A. Yildirim, K.-I. Goh, M. E. Cusick, A.-L. Barabási, and M. Vidal, "Drug-target network.," *Nature biotechnology*, vol. 25, pp. 1119–26, Oct. 2007.

[127] A. Pujol, R. Mosca, J. Farrés, and P. Aloy, "Unveiling the role of network and systems biology in drug discovery.," *Trends in pharmacological sciences*, vol. 31, pp. 115–23, Mar. 2010.

[128] D. S. Wishart, "Chapter 3: Small molecules and disease.," *PLoS computational biology*, vol. 8, p. e1002805, Jan. 2012.

[129] A. P. Davis, T. C. Wiegers, M. C. Rosenstein, and C. J. Mattingly, "MEDIC: a practical disease vocabulary used at the Comparative Toxicogenomics Database.," *Database : the journal of biological databases and curation*, vol. 2012, p. bar065, Jan. 2012.

[130] M. D. Shultz, "Setting expectations in molecular optimizations: Strengths and limitations of commonly used composite parameters.," *Bioorganic & medicinal chemistry letters*, vol. 23, pp. 5980–91, Nov. 2013.

[131] M. Congreve, R. Carr, C. Murray, and H. Jhoti, "A 'rule of three' for fragment-based lead discovery?," *Drug discovery today*, vol. 8, pp. 876–7, Oct. 2003.

[132] S. R. Langdon, N. Brown, and J. Blagg, "Scaffold diversity of exemplified medicinal chemistry space.," *Journal of chemical information and modeling*, vol. 51, pp. 2174–85, Sept. 2011.

[133] I. Akritopoulou-Zanze and P. J. Hajduk, "Kinase-targeted libraries: the design and synthesis of novel, potent, and selective kinase inhibitors.," *Drug discovery today*, vol. 14, pp. 291–7, Mar. 2009.

[134] E.-W. Lameijer, J. N. Kok, T. Bäck, and A. P. Ijzerman, "Mining a chemical database for fragment co-occurrence: discovery of "chemical clichés".," *Journal of chemical information and modeling*, vol. 46, pp. 553–62, Jan. 2006.

[135] A. H. Lipkus, Q. Yuan, K. A. Lucas, S. A. Funk, and e. a. Bartelt, "Structural diversity of organic chemistry. A scaffold analysis of the CAS Registry.," *The Journal of organic chemistry*, vol. 73, pp. 4443–51, June 2008.

[136] W. R. Pitt, D. M. Parry, B. G. Perry, and C. R. Groom, "Heteroaromatic rings of the future.," *Journal of medicinal chemistry*, vol. 52, pp. 2952–63, May 2009.

[137] D. P. Williams, "Toxicophores: investigations in drug safety.," *Toxicology*, vol. 226, pp. 1–11, Sept. 2006.

[138] M. E. Welsch, S. A. Snyder, and B. R. Stockwell, "Privileged scaffolds for library design and drug discovery.," *Current opinion in chemical biology*, vol. 14, pp. 347–61, June 2010.

[139] M. J. Wester, S. N. Pollock, E. A. Coutsias, T. K. Allu, and e. a. Muresan, "Scaffold topologies. 2. Analysis of chemical databases.," *Journal of chemical information and modeling*, vol. 48, pp. 1311–24, July 2008.

[140] E. Dupont, B. Mikkelsen, and J. Jakobsen, "Mesulergine in early Parkinson's disease: a double blind controlled trial.," *Journal of neurology, neurosurgery, and psychiatry*, vol. 49, pp. 390–5, Apr. 1986.

[141] L. Mavridis, B. D. Hudson, and D. W. Ritchie, "Toward high throughput 3D virtual screening using spherical harmonic surface representations.," *Journal of chemical information and modeling*, vol. 47, pp. 1787–96, Jan. 2007.

[142] I. Takigawa and H. Mamitsuka, "Graph mining: procedure, application to drug discovery and recent advances.," *Drug discovery today*, vol. 18, pp. 50–7, Jan. 2013.

[143] Y. Wang, J. Lu, F. Wang, Q. Shen, and e. a. Zheng, "Estimation of carcinogenicity using molecular fragments tree.," *Journal of chemical information and modeling*, vol. 52, pp. 1994–2003, Aug. 2012.

[144] N. Greene, P. N. Judson, J. J. Langowski, and C. A. Marchant, "Knowledge-based expert systems for toxicity and metabolism prediction: DEREK, StAR and METEOR.," *SAR and QSAR in environmental research*, vol. 10, pp. 299–314, Jan. 1999.

[145] S. I. Berger, A. Ma'ayan, and R. Iyengar, "Systems pharmacology of arrhythmias.," *Science signaling*, vol. 3, p. ra30, Jan. 2010.

[146] A. Nakajima, K. Doki, M. Homma, T. Sagae, and e. a. Saito, "Investigation of glucocorticoid-induced side effects in patients with autoimmune diseases.," *Yakugaku zasshi : Journal of the Pharmaceutical Society of Japan*, vol. 129, pp. 445–50, Apr. 2009.

[147] P. Yang, X.-L. Li, J.-P. Mei, C.-K. Kwoh, and e. a. Ng, "Positive-unlabeled learning for disease gene identification.," *Bioinformatics (Oxford, England)*, vol. 28, pp. 2640–7, Oct. 2012.

[148] B. Zhang and W. Zuo, "Reliable Negative Extracting Based on kNN for Learning from Positive and Unlabeled Examples," *Journal of Computers*, vol. 4, pp. 94–101, Jan. 2009.

[149] B. Liu, W. S. Lee, P. S. Yu, and X. Li, "Partially Supervised Classification of Text Documents," pp. 387–394, July 2002.

[150] A. P. Davis, T. C. Wiegers, P. M. Roberts, B. L. King, and e. a. Lay, "A CTD-Pfizer collaboration: manual curation of 88,000 scientific articles text mined for drug-disease and drug-phenotype interactions.," *Database : the journal of biological databases and curation*, vol. 2013, p. bat080, Jan. 2013.

[151] C. Y. Liew, C. Pan, A. Tan, K. X. M. Ang, and e. a. Yap, "QSAR classification of metabolic activation of chemicals into covalently reactive species.," *Molecular diversity*, vol. 16, pp. 389–400, May 2012.

[152] M. Casalegno and G. Sello, "Determination of toxicant mode of action by augmented top priority fragment class.," *Journal of chemical information and modeling*, vol. 53, pp. 1113–26, May 2013.

[153] A. Jacunski and N. P. Tatonetti, "Connecting the dots: applications of network medicine in pharmacology and disease.," *Clinical pharmacology and therapeutics*, vol. 94, pp. 659–69, Dec. 2013.

[154] K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, and e. a. Vidal, "The human disease network.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, pp. 8685–90, May 2007.

[155] C. A. Hidalgo, N. Blumm, A.-L. Barabási, and N. A. Christakis, "A dynamic network approach for the study of human phenotypes.," *PLoS computational biology*, vol. 5, p. e1000353, Apr. 2009.

[156] D.-S. Lee, J. Park, K. A. Kay, N. A. Christakis, and e. a. Oltvai, "The impli-cations of human metabolic network topology for disease comorbidity.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, pp. 9880–5, July 2008.

[157] T. T. Ashburn and K. B. Thor, "Drug repositioning: identifying and devel-oping new uses for existing drugs.," *Nature reviews. Drug discovery*, vol. 3, pp. 673–83, Aug. 2004.

[158] A. P. Chiang and A. J. Butte, "Systematic evaluation of drug-disease rela-tionships to identify leads for novel drug uses.," *Clinical pharmacology and therapeutics*, vol. 86, pp. 507–10, Nov. 2009.

[159] W. Gielen, T. J. Cleophas, and R. Agrawal, "Nebivolol: a review of its clin-ical and pharmacological characteristics.," *International journal of clinical pharmacology and therapeutics*, vol. 44, pp. 344–57, Aug. 2006.

[160] M. Duran-Frigola and P. Aloy, "Recycling side-effects into clinical mark-ers for drug repositioning.," *Genome medicine*, vol. 4, p. 3, Jan. 2012.

[161] R. B. Russell and P. Aloy, "Targeting and tinkering with interaction net-works.," *Nature chemical biology*, vol. 4, pp. 666–73, Nov. 2008.

[162] K. Audouze, A. S. Juncker, F. J. S. S. A. Roque, K. Krysiak-Baltyn, and e. a. Weinhold, "Deciphering diseases and biological targets for environ-mental chemicals using toxicogenomics networks.," *PLoS computational biology*, vol. 6, p. e1000788, May 2010.

[163] T. Liu, Y. Lin, X. Wen, R. N. Jorissen, and e. a. Gilson, "BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities.," *Nucleic acids research*, vol. 35, pp. D198–201, Jan. 2007.

[164] A. P. Davis, T. C. Wiegers, R. J. Johnson, J. M. Lay, and e. a. Lennon-Hopkins, "Text mining effectively scores and ranks the literature for improving chemical-gene-disease curation at the comparative toxicoge-nomics database.," *PloS one*, vol. 8, p. e58201, Jan. 2013.

[165] T. Grego and F. M. Couto, "Enhancement of chemical entity identification in text using semantic similarity validation.," *PloS one*, vol. 8, p. e62984, Jan. 2013.

[166] Y. Ding, M. Song, J. Han, Q. Yu, and e. a. Yan, "Entitymetrics: measuring the impact of entities.," *PloS one*, vol. 8, p. e71416, Jan. 2013.

[167] Z. Hu, Y.-C. Chang, Y. Wang, C.-L. Huang, and e. a. Liu, "VisANT 4.0: Integrative network platform to connect genes, drugs, diseases and ther-apies.," *Nucleic acids research*, vol. 41, pp. W225–31, July 2013.

[168] T. I. Oprea, E. E. May, A. Leitão, and A. Tropsha, "Computational systems chemical biology.," *Methods in molecular biology (Clifton, N.J.)*, vol. 672, pp. 459–88, Jan. 2011.

[169] B. L. King, A. P. Davis, M. C. Rosenstein, T. C. Wiegers, and e. a. Mattingly, "Ranking transitive chemical-disease inferences using local network topology in the comparative toxicogenomics database.," *PloS one*, vol. 7, p. e46524, Jan. 2012.

[170] D. Butina, "Unsupervised Data Base Clustering Based on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way To Cluster Small and Large Data Sets," *Journal of Chemical Information and Modeling*, vol. 39, pp. 747–750, July 1999.

[171] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," pp. 321–357, June 2011.

[172] D. Müllner, "fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python," *Journal of Statistical Software*, vol. 53, no. 9, pp. 1–18, 2013.

[173] P. Langfelder, B. Zhang, and S. Horvath, "Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R.," *Bioinformatics (Oxford, England)*, vol. 24, pp. 719–20, Mar. 2008.

[174] J. Wildenhain, N. Fitzgerald, and M. Tyers, "MolClass: a web portal to interrogate diverse small molecule screen datasets with different computational models.," *Bioinformatics (Oxford, England)*, vol. 28, pp. 2200–1, Aug. 2012.

[175] A. Liaw and M. Wiener, "Classification and Regression by randomForest.," *R News*, vol. 2, pp. 18–22, 2002.

[176] B. T. McInnes, T. Pedersen, and S. V. S. Pakhomov, "UMLS-Interface and UMLS-Similarity : open source software for measuring paths and semantic similarity.," *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, vol. 2009, pp. 431–5, Jan. 2009.

[177] L. Yang and P. Agarwal, "Systematic drug repositioning based on clinical side-effects.," *PloS one*, vol. 6, p. e28025, Jan. 2011.

[178] S. Jaeger, M. Duran-Frigola, and P. Aloy, "Drug sensitivity in cancer cell lines is not tissue-specific.," *Molecular cancer*, vol. 14, p. 40, Jan. 2015.

[179] J. N. Weinstein, "Drug discovery: Cell lines battle cancer.," *Nature*, vol. 483, pp. 544–5, Mar. 2012.

[180] M. J. Garnett, E. J. Edelman, S. J. Heidorn, C. D. Greenman, and e. a. Dastur, "Systematic identification of genomic markers of drug sensitivity in cancer cells.," *Nature*, vol. 483, pp. 570–5, Mar. 2012.

[181] A. Prahallad, C. Sun, S. Huang, F. Di Nicolantonio, R. Salazar, D. Zecchin, R. L. Beijersbergen, A. Bardelli, and R. Bernards, "Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR.," *Nature*, vol. 483, pp. 100–3, Mar. 2012.

[182] A. F. Gazdar, L. Girard, W. W. Lockwood, W. L. Lam, and e. a. Minna, "Lung cancer cell lines as tools for biomedical discovery and research.," *Journal of the National Cancer Institute*, vol. 102, pp. 1310–21, Sept. 2010.

[183] R. H. Shoemaker, "The NCI60 human tumour cell line anticancer drug screen.," *Nature reviews. Cancer*, vol. 6, pp. 813–23, Oct. 2006.

[184] A. F. Gazdar, B. Gao, and J. D. Minna, "Lung cancer cell lines: Useless artifacts or invaluable tools for medical science?," *Lung cancer (Amsterdam, Netherlands)*, vol. 68, pp. 309–18, June 2010.

[185] J.-P. Gillet, S. Varma, and M. M. Gottesman, "The clinical relevance of cancer cell lines.," *Journal of the National Cancer Institute*, vol. 105, pp. 452–8, Apr. 2013.

[186] S. Domcke, R. Sinha, D. A. Levine, C. Sander, and e. a. Schultz, "Evaluating cell lines as tumour models by comparison of genomic profiles.," *Nature communications*, vol. 4, p. 2126, Jan. 2013.

[187] J. L. Wilding and W. F. Bodmer, "Cancer cell lines for drug discovery and development.," *Cancer research*, vol. 74, pp. 2377–84, May 2014.

[188] A. Marusyk and K. Polyak, "Tumor heterogeneity: causes and consequences.," *Biochimica et biophysica acta*, vol. 1805, pp. 105–17, Jan. 2010.

[189] T. Vargo-Gogola and J. M. Rosen, "Modelling breast cancer: one size does not fit all.," *Nature reviews. Cancer*, vol. 7, pp. 659–72, Sept. 2007.

[190] C. M. Perou, T. Sø rlie, M. B. Eisen, M. van de Rijn, and e. a. Jeffrey, "Molecular portraits of human breast tumours.," *Nature*, vol. 406, pp. 747–52, Aug. 2000.

[191] L. J. van 't Veer, H. Dai, M. J. van de Vijver, Y. D. He, A. A. M. Hart, M. Mao, H. L. Peterse, K. van der Kooy, M. J. Marton, A. T. Witteveen, G. J. Schreiber, R. M. Kerkhoven, C. Roberts, P. S. Linsley, R. Bernards, and S. H. Friend, "Gene expression profiling predicts clinical outcome of breast cancer.," *Nature*, vol. 415, pp. 530–6, Jan. 2002.

[192] D. L. Holliday and V. Speirs, "Choosing the right cell line for breast cancer research.," *Breast cancer research : BCR*, vol. 13, p. 215, Jan. 2011.

[193] J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, and e. a. Margolin, "The Cancer Cell Line Encyclopedia enables predictive modelling of anti-cancer drug sensitivity.," *Nature*, vol. 483, pp. 603–7, Mar. 2012.

[194] A. Moghaddas Gholami, H. Hahne, Z. Wu, F. J. Auer, C. Meng, M. Wilhelm, and B. Kuster, "Global proteome analysis of the NCI-60 cell line panel.," *Cell reports*, vol. 4, pp. 609–20, Aug. 2013.

[195] F. A. Daniela Ferreira and R. Chaves, *Oncogenomics and Cancer Proteomics - Novel Approaches in Biomarkers Discovery and Therapeutic Targets in Cancer*. InTech, Mar. 2013.

[196] O. D. Abaan, E. C. Polley, S. R. Davis, Y. J. Zhu, and e. a. Bilke, "The exomes of the NCI-60 panel: a genomic resource for cancer biology and systems pharmacology.," *Cancer research*, vol. 73, pp. 4372–82, July 2013.

[197] D. T. Ross, U. Scherf, M. B. Eisen, C. M. Perou, and e. a. Rees, "Systematic variation in gene expression patterns in human cancer cell lines.," *Nature genetics*, vol. 24, pp. 227–35, Mar. 2000.

[198] G. M. Dancik, Y. Ru, C. R. Owens, and D. Theodorescu, "A framework to select clinically relevant cancer cell lines for investigation by establishing their molecular similarity with primary human cancers.," *Cancer research*, vol. 71, pp. 7398–409, Dec. 2011.

[199] C. Qin, C. Zhang, F. Zhu, F. Xu, and e. a. Chen, "Therapeutic target database update 2014: a resource for targeted therapeutics.," *Nucleic acids research*, vol. 42, pp. D1118–23, Jan. 2014.

[200] B. Haibe-Kains, N. El-Hachem, N. J. Birkbak, A. C. Jin, and e. a. Beck, "Inconsistency in large pharmacogenomic studies.," *Nature*, vol. 504, pp. 389–93, Dec. 2013.

[201] M. G. Vander Heiden, L. C. Cantley, and C. B. Thompson, "Understanding the Warburg effect: the metabolic requirements of cell proliferation.," *Science (New York, N.Y.)*, vol. 324, pp. 1029–33, May 2009.

[202] A. Schulze and A. L. Harris, "How cancer metabolism is tuned for proliferation and vulnerable to disruption.," *Nature*, vol. 491, pp. 364–73, Nov. 2012.

[203] B. A. Webb, M. Chimenti, M. P. Jacobson, and D. L. Barber, "Dysregulated pH: a perfect storm for cancer progression.," *Nature reviews. Cancer*, vol. 11, pp. 671–7, Sept. 2011.

[204] J. R. Casey, S. Grinstein, and J. Orlowski, "Sensors and regulators of intracellular pH.," *Nature reviews. Molecular cell biology*, vol. 11, pp. 50–61, Jan. 2010.

[205] G. Helmlinger, F. Yuan, M. Dellian, and R. K. Jain, "Interstitial pH and pO2 gradients in solid tumors in vivo: high-resolution measurements reveal a lack of correlation.," *Nature medicine*, vol. 3, pp. 177–82, Feb. 1997.

[206] J. Chiche, M. C. Brahimi-Horn, and J. Pouysségur, "Tumour hypoxia induces a metabolic shift causing acidosis: a common feature in cancer.," *Journal of cellular and molecular medicine*, vol. 14, pp. 771–94, Apr. 2010.

[207] R. A. Gatenby and R. J. Gillies, "Why do cancers have high aerobic glycolysis?," *Nature reviews. Cancer*, vol. 4, pp. 891–9, Nov. 2004.

[208] R. A. Gatenby, E. T. Gawlinski, A. F. Gmitro, B. Kaylor, and e. a. Gillies, "Acid-mediated tumor invasion: a multidisciplinary study.," *Cancer research*, vol. 66, pp. 5216–23, May 2006.

[209] I. F. Robey, B. K. Baggett, N. D. Kirkpatrick, D. J. Roe, and e. a. Dosescu, "Bicarbonate increases tumor pH and inhibits spontaneous metastases.," *Cancer research*, vol. 69, pp. 2260–8, Mar. 2009.

[210] M. F. McCarty and J. Whitaker, "Manipulating tumor acidification as a cancer treatment strategy.," *Alternative medicine review : a journal of clinical therapeutic*, vol. 15, pp. 264–72, Sept. 2010.

[211] R. J. Gillies, D. Verduzco, and R. A. Gatenby, "Evolutionary dynamics of carcinogenesis and why targeted therapy does not work.," *Nature reviews. Cancer*, vol. 12, pp. 487–93, July 2012.

[212] R. J. Gillies and R. A. Gatenby, "Metabolism and its sequelae in cancer evolution and therapy.," *Cancer journal (Sudbury, Mass.)*, vol. 21, pp. 88–96, Jan. 2015.

[213] I. F. Tannock and D. Rotin, "Acid pH in tumors and its potential for therapeutic exploitation.," *Cancer research*, vol. 49, pp. 4373–84, Aug. 1989.

[214] L. E. Gerweck and K. Seetharaman, "Cellular pH gradient in tumor versus normal tissue: potential exploitation for the treatment of cancer.," *Cancer research*, vol. 56, pp. 1194–8, Mar. 1996.

[215] D. Neri and C. T. Supuran, "Interfering with pH regulation in tumours as a therapeutic strategy.," *Nature reviews. Drug discovery*, vol. 10, pp. 767–77, Oct. 2011.

[216] B. K. Grillo-Hill, C. Choi, M. Jimenez-Vidal, and D. L. Barber, "Increased H+ efflux is sufficient to induce dysplasia and necessary for viability with oncogene expression.," *eLife*, vol. 4, Jan. 2015.

[217] T. B. Dansen, K. W. Wirtz, R. J. Wanders, and E. H. Pap, "Peroxisomes in human fibroblasts have a basic pH.," *Nature cell biology*, vol. 2, pp. 51–3, Jan. 2000.

[218] J. R. Doherty, C. Yang, K. E. N. Scott, M. D. Cameron, and e. a. Fallahi, "Blocking lactate export by inhibiting the Myc target MCT1 Disables glycolysis and glutathione synthesis.," *Cancer research*, vol. 74, pp. 908–20, Feb. 2014.

[219] K. Yizhak, S. E. Le Dévédec, V. M. Rogkoti, F. Baenke, V. C. de Boer, C. Frezza, A. Schulze, B. van de Water, and E. Ruppin, "A computational study of the Warburg effect identifies metabolic targets inhibiting cancer migration.," *Molecular systems biology*, vol. 10, p. 744, Jan. 2014.

[220] N. D. Price, J. L. Reed, and B. O. Palsson, "Genome-scale models of microbial cells: evaluating the consequences of constraints.," *Nature reviews. Microbiology*, vol. 2, pp. 886–97, Nov. 2004.

[221] B. Garcia-Moreno, "Adaptations of proteins to cellular and subcellular pH.," *Journal of biology*, vol. 8, p. 98, Jan. 2009.

[222] E. Alexov, "Numerical calculations of the pH of maximal protein stability. The effect of the sequence composition and three-dimensional structure.," *European journal of biochemistry / FEBS*, vol. 271, pp. 173–85, Jan. 2004.

[223] K. Talley and E. Alexov, "On the pH-optimum of activity and stability of proteins.," *Proteins*, vol. 78, pp. 2699–706, Sept. 2010.

[224] V. Kunik, Y. Meroz, Z. Solan, B. Sandbank, and e. a. Weingart, "Functional representation of enzymes by specific peptides.," *PLoS computational biology*, vol. 3, p. e167, Aug. 2007.

[225] W. Tian, A. K. Arakaki, and J. Skolnick, "EFICAz: a comprehensive approach for accurate genome-scale enzyme function inference.," *Nucleic acids research*, vol. 32, pp. 6226–39, Jan. 2004.

[226] M. Duran-Frigola, R. Mosca, and P. Aloy, "Structural Systems Pharmacology: The Role of 3D Structures in Next-Generation Drug Development," *Chemistry & Biology*, vol. 20, pp. 674–684, May 2013.

[227] G. V. Paolini, R. H. B. Shapland, W. P. van Hoorn, J. S. Mason, and e. a. Hopkins, "Global mapping of pharmacological space.," *Nature biotechnology*, vol. 24, pp. 805–15, July 2006.

[228] L. Cucurull-Sanchez, K. G. Spink, and S. A. Moschos, "Relevance of systems pharmacology in drug discovery.," *Drug discovery today*, vol. 17, pp. 665–70, July 2012.

[229] J. N. Y. Chan, C. Nislow, and A. Emili, "Recent advances and method development for drug target identification.," *Trends in pharmacological sciences*, vol. 31, pp. 82–8, Feb. 2010.

[230] M. Iskar, G. Zeller, X.-M. Zhao, V. van Noort, and e. a. Bork, "Drug discovery in the age of systems biology: the rise of computational approaches for data integration," *Current Opinion in Biotechnology*, vol. 23, pp. 609–616, Aug. 2012.

[231] C. Laggner, D. Kokel, V. Setola, A. Tolia, and e. a. Lin, "Chemical informatics and target identification in a zebrafish phenotypic screen.," *Nature chemical biology*, vol. 8, pp. 144–6, Feb. 2012.

[232] E. Gregori-Puigjané, V. Setola, J. Hert, B. A. Crews, J. J. Irwin, E. Lounkine, L. Marnett, B. L. Roth, and B. K. Shoichet, "Identifying mechanism-of-action targets for drugs and probes.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, pp. 11178–83, July 2012.

[233] P. Aloy and R. B. Russell, "Structure-based systems biology: a zoom lens for the cell.," *FEBS letters*, vol. 579, pp. 1854–8, Mar. 2005.

[234] J. B. Brown and Y. Okuno, "Systems biology and systems chemistry: new directions for drug discovery.," *Chemistry & biology*, vol. 19, pp. 23–8, Jan. 2012.

[235] P. Aloy and R. B. Russell, "Structural systems biology: modelling protein interactions.," *Nature reviews. Molecular cell biology*, vol. 7, pp. 188–97, Mar. 2006.

[236] G. Schneider and U. Fechner, "Computer-based de novo design of drug-like molecules.," *Nature reviews. Drug discovery*, vol. 4, pp. 649–63, Aug. 2005.

[237] A.-L. Barabási and Z. N. Oltvai, "Network biology: understanding the cell's functional organization.," *Nature reviews. Genetics*, vol. 5, pp. 101–13, Feb. 2004.

[238] T. Yamada and P. Bork, "Evolution of biomolecular networks: lessons from metabolic and protein interactions.," *Nature reviews. Molecular cell biology*, vol. 10, pp. 791–803, Nov. 2009.

[239] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, and e. a. Bhat, "The Protein Data Bank.," *Nucleic acids research*, vol. 28, pp. 235–42, Jan. 2000.

[240] G. R. Bickerton, A. P. Higueruelo, and T. L. Blundell, "Comprehensive, atomic-level characterization of structurally characterized protein-protein interactions: the PICCOLO database.," *BMC bioinformatics*, vol. 12, p. 313, Jan. 2011.

[241] A. Stein and P. Aloy, "Novel peptide-mediated interactions derived from high-resolution 3-dimensional structures.," *PLoS computational biology*, vol. 6, p. e1000789, May 2010.

[242] A. Stein, R. Mosca, and P. Aloy, "Three-dimensional modeling of protein interactions and complexes is going 'omics.," *Current opinion in structural biology*, vol. 21, pp. 200–8, Apr. 2011.

[243] C. Winter, A. Henschel, A. Tuukkanen, and M. Schroeder, "Protein interactions in 3D: from interface evolution to drug discovery.," *Journal of structural biology*, vol. 179, pp. 347–58, Sept. 2012.

[244] X. Wang, X. Wei, B. Thijssen, J. Das, and e. a. Lipkin, "Three-dimensional reconstruction of protein networks provides insight into human genetic disease.," *Nature biotechnology*, vol. 30, pp. 159–64, Feb. 2012.

[245] M. Gao and J. Skolnick, "Structural space of protein-protein interfaces is degenerate, close to complete, and highly connected.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, pp. 22517–22, Dec. 2010.

[246] P. J. Kundrotas, Z. Zhu, J. Janin, and I. A. Vakser, "Templates are available to model nearly all complexes of structurally characterized proteins.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, pp. 9438–41, June 2012.

[247] R. Mosca, C. Pons, J. Fernández-Recio, and P. Aloy, "Pushing structural information into the yeast interactome by high-throughput protein docking experiments.," *PLoS computational biology*, vol. 5, p. e1000490, Aug. 2009.

[248] Q. C. Zhang, D. Petrey, L. Deng, L. Qiang, and e. a. Shi, "Structure-based prediction of protein-protein interactions on a genome-wide scale.," *Nature*, vol. 490, pp. 556–60, Oct. 2012.

[249] C. N. Cavasotto and S. S. Phatak, "Homology modeling in drug discovery: current trends and applications.," *Drug discovery today*, vol. 14, pp. 676–83, July 2009.

[250] J. Skolnick, H. Zhou, and M. Gao, "Are predicted protein structures of any value for binding site prediction and virtual ligand screening?," *Current opinion in structural biology*, vol. 23, pp. 191–7, Apr. 2013.

[251] G. Kar, A. Gursoy, and O. Keskin, "Human cancer protein-protein interaction network: a structural perspective.," *PLoS computational biology*, vol. 5, p. e1000601, Dec. 2009.

[252] G. Kuzu, O. Keskin, A. Gursoy, and R. Nussinov, "Constructing structural networks of signaling pathways on the proteome scale.," *Current opinion in structural biology*, vol. 22, pp. 367–77, June 2012.

[253] Y. Zhang, I. Thiele, D. Weekes, Z. Li, and e. a. Jaroszewski, "Three-dimensional structural view of the central metabolic network of Thermotoga maritima.," *Science (New York, N.Y.)*, vol. 325, pp. 1544–9, Sept. 2009.

[254] S. L. Kinnings, L. Xie, K. H. Fung, R. M. Jackson, and e. a. Xie, "The Mycobacterium tuberculosis drugome and its polypharmacological implications.," *PLoS computational biology*, vol. 6, p. e1000976, Jan. 2010.

[255] J. von Eichborn, M. Dunkel, B. O. Gohlke, S. C. Preissner, M. F. Hoffmann, J. M. J. Bauer, J. D. Armstrong, M. H. Schaefer, M. A. Andrade-Navarro, N. Le Novere, M. D. R. Croning, S. G. N. Grant, P. van Nierop, A. B. Smit, and R. Preissner, "SynSysNet: integration of experimental data on synaptic protein-protein interactions with drug-target relations.," *Nucleic acids research*, vol. 41, pp. D834–40, Jan. 2013.

[256] A. L. Hopkins, "Network pharmacology: the next paradigm in drug discovery.," *Nature chemical biology*, vol. 4, pp. 682–90, Nov. 2008.

[257] B. P. Zambrowicz and A. T. Sands, "Knockouts model the 100 best-selling drugs–will they model the next 100?," *Nature reviews. Drug discovery*, vol. 2, pp. 38–51, Jan. 2003.

[258] V. Agoston, P. Csermely, and S. Pongor, "Multiple weak hits confuse complex systems: a transcriptional regulatory network as an example.," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 71, p. 051909, May 2005.

[259] G. E. Winter, U. Rix, S. M. Carlson, K. V. Gleixner, and e. a. Gre-bien, "Systems-pharmacology dissection of a drug synergy in imatinib-resistant CML.," *Nature chemical biology*, vol. 8, pp. 905–12, Nov. 2012.

[260] H. Kitano, "A robustness-based approach to systems-oriented drug de-sign.," *Nature reviews. Drug discovery*, vol. 6, pp. 202–10, Mar. 2007.

[261] K. Yang, H. Bai, Q. Ouyang, L. Lai, and e. a. Tang, "Finding multiple tar-get optimal intervention in disease-related molecular network.," *Molecu-lar systems biology*, vol. 4, p. 228, Jan. 2008.

[262] N. M. Penrod, R. Cowper-Sal-lari, and J. H. Moore, "Systems genetics for drug target discovery.," *Trends in pharmacological sciences*, vol. 32, pp. 623–30, Oct. 2011.

[263] W.-C. Hwang, A. Zhang, and M. Ramanathan, "Identification of informa-tion flow-modulating drug targets: a novel bridging paradigm for drug discovery.," *Clinical pharmacology and therapeutics*, vol. 84, pp. 563–72, Nov. 2008.

[264] H. Yu, P. M. Kim, E. Sprecher, V. Trifonov, and e. a. Gerstein, "The impor-tance of bottlenecks in protein networks: correlation with gene essential-ity and expression dynamics.," *PLoS computational biology*, vol. 3, p. e59, Apr. 2007.

[265] J. Wang, Z.-x. Li, C.-x. Qiu, D. Wang, and e. a. Cui, "The relationship between rational drug design and drug side effects.," *Briefings in bioinfor-matics*, vol. 13, pp. 377–82, May 2012.

[266] Q.-W. Fan, Z. A. Knight, D. D. Goldenberg, W. Yu, and e. a. Mostov, "A dual PI3 kinase/mTOR inhibitor reveals emergent efficacy in glioma.," *Cancer cell*, vol. 9, pp. 341–9, May 2006.

[267] Z. A. Knight, H. Lin, and K. M. Shokat, "Targeting the cancer kinome through polypharmacology.," *Nature reviews. Cancer*, vol. 10, pp. 130–7, Feb. 2010.

[268] S. Gordo, V. Martos, E. Santos, M. Menéndez, C. Bo, E. Giralt, and J. de Mendoza, "Stability and structural recovery of the tetramerization domain of p53-R337H mutant induced by a designed templating ligand.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, pp. 16426–31, Oct. 2008.

[269] C. E. Schafmeister, J. Po, and G. L. Verdine, "An All-Hydrocarbon Cross-Linking System for Enhancing the Helicity and Metabolic Stability of Pep-tides," *Journal of the American Chemical Society*, vol. 122, pp. 5891–5892, June 2000.

[270] E. F. Lee, P. E. Czabotar, B. J. Smith, K. Deshayes, and e. a. Zobel, "Crystal structure of ABT-737 complexed with Bcl-xL: implications for selectivity of antagonists of the Bcl-2 family.," *Cell death and differentiation*, vol. 14, pp. 1711–3, Sept. 2007.

[271] D. Kozakov, D. R. Hall, G.-Y. Chuang, R. Cencic, and e. a. Brenke, "Structural conservation of druggable hot spots in protein-protein interfaces.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, pp. 13528–33, Aug. 2011.

[272] S. Eyrisch and V. Helms, "Transient pockets on protein surfaces involved in protein-protein interaction.," *Journal of medicinal chemistry*, vol. 50, pp. 3457–64, July 2007.

[273] S. H. Lawrence, U. D. Ramirez, L. Tang, F. Fazliyez, and e. a. Kundrat, "Shape shifting leads to small-molecule allosteric drug discovery.," *Chemistry & biology*, vol. 15, pp. 586–96, June 2008.

[274] J.-J. Lu, W. Pan, Y.-J. Hu, and Y.-T. Wang, "Multi-target drugs: the trend of drug research and development.," *PloS one*, vol. 7, p. e40262, Jan. 2012.

[275] D. Wei, X. Jiang, L. Zhou, J. Chen, and e. a. Chen, "Discovery of multi-target inhibitors by combining molecular docking with common pharmacophore matching.," *Journal of medicinal chemistry*, vol. 51, pp. 7882–8, Dec. 2008.

[276] S. L. Schreiber, "Small molecules: the missing link in the central dogma.," *Nature chemical biology*, vol. 1, pp. 64–6, July 2005.

[277] A. M. Edwards, C. Bountra, D. J. Kerr, and T. M. Willson, "Open access chemical and clinical probes to support drug discovery.," *Nature chemical biology*, vol. 5, pp. 436–40, July 2009.

[278] A. P. Russ and S. Lampel, "The druggable genome: an update.," *Drug discovery today*, vol. 10, pp. 1607–10, Dec. 2005.

[279] J. D. Durrant, R. E. Amaro, L. Xie, M. D. Urbaniak, and e. a. Ferguson, "A multidimensional strategy to detect polypharmacological targets in the absence of structural and sequence homology.," *PLoS computational biology*, vol. 6, p. e1000648, Jan. 2010.

[280] J.-I. Ito, Y. Tabei, K. Shimizu, K. Tsuda, and e. a. Tomii, "PoSSuM: a database of similar protein-ligand binding and putative pockets.," *Nucleic acids research*, vol. 40, pp. D541–8, Jan. 2012.

[281] I. Kufareva, A. V. Ilatovskiy, and R. Abagyan, "Pocketome: an encyclopedia of small-molecule binding sites in 4D.," *Nucleic acids research*, vol. 40, pp. D535–40, Jan. 2012.

[282] T. Beuming and W. Sherman, "Current assessment of docking into GPCR crystal structures and homology models: successes, challenges, and guidelines.," *Journal of chemical information and modeling*, vol. 52, pp. 3263–77, Dec. 2012.

[283] M. Brylinski and J. Skolnick, "Cross-reactivity virtual profiling of the human kinome by X-react(KIN): a chemical systems biology approach.," *Molecular pharmaceutics*, vol. 7, pp. 2324–33, Dec. 2010.

[284] J.-C. Wang, P.-Y. Chu, C.-M. Chen, and J.-H. Lin, "idTarget: a web server for identifying protein targets of small chemical molecules with robust scoring functions and a divide-and-conquer docking approach.," *Nucleic acids research*, vol. 40, pp. W393–9, July 2012.

[285] L. Xie, L. Xie, and P. E. Bourne, "Structure-based systems biology for analyzing off-target binding.," *Current opinion in structural biology*, vol. 21, pp. 189–99, Apr. 2011.

[286] F. Cheng, C. Liu, J. Jiang, W. Lu, and e. a. Li, "Prediction of drug-target interactions and drug repositioning via network-based inference.," *PLoS computational biology*, vol. 8, p. e1002503, Jan. 2012.

[287] P. Workman and I. Collins, "Probing the probes: fitness factors for small molecule tools.," *Chemistry & biology*, vol. 17, pp. 561–77, June 2010.

[288] O. Taboureau, J. B. Baell, J. Fernández-Recio, and B. O. Villoutreix, "Established and emerging trends in computational drug discovery in the structural genomics era.," *Chemistry & biology*, vol. 19, pp. 29–41, Jan. 2012.

[289] Y. Shen, J. Liu, G. Estiu, B. Isin, and e. a. Ahn, "Blueprint for antimicrobial hit discovery targeting metabolic networks.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, pp. 1082–7, Jan. 2010.

[290] O. Folger, L. Jerby, C. Frezza, E. Gottlieb, and e. a. Ruppin, "Predicting selective drug targets in cancer through metabolic networks.," *Molecular systems biology*, vol. 7, p. 501, Jan. 2011.

[291] S. Henrich, O. M. H. Salo-Ahen, B. Huang, F. F. Rippmann, and e. a. Cruciani, "Computational approaches to identifying and characterizing protein binding sites for ligand design.," *Journal of molecular recognition*, vol. 23, pp. 209–19, Jan. 2009.

[292] J. S. Mason, A. Bortolato, M. Congreve, and F. H. Marshall, "New insights from structural biology into the druggability of G protein-coupled receptors.," *Trends in pharmacological sciences*, vol. 33, pp. 249–60, May 2012.

[293] Q. Zhong, N. Simonis, Q.-R. Li, B. Charloteaux, and e. a. Heuze, "Edgetic perturbation models of human inherited disorders," *Molecular Systems Biology*, vol. 5, p. 321, Nov. 2009.

[294] M. J. Basse, S. Betzi, R. Bourgeas, S. Bouzidi, B. Chetrit, V. Hamon, X. Morelli, and P. Roche, "2P2Idb: a structural database dedicated to orthosteric modulation of protein-protein interactions.," *Nucleic acids research*, vol. 41, pp. D824–7, Jan. 2013.

[295] A. Mullard, "Protein-protein interaction inhibitors get into the groove.," *Nature reviews. Drug discovery*, vol. 11, pp. 173–5, Mar. 2012.

[296] J. C. Fuller, N. J. Burgoyne, and R. M. Jackson, "Predicting druggable binding sites at the protein-protein interface.," *Drug discovery today*, vol. 14, pp. 155–61, Mar. 2009.

[297] U. Schlecht, M. Miranda, S. Suresh, R. W. Davis, and R. P. St Onge, "Multiplex assay for condition-dependent changes in protein-protein interactions.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, pp. 9213–8, June 2012.

[298] S. Eyrisch and V. Helms, "What induces pocket openings on protein surface patches involved in protein-protein interactions?," *Journal of computer-aided molecular design*, vol. 23, pp. 73–86, Feb. 2009.

[299] A. Metz, C. Pfleger, H. Kopitz, S. Pfeiffer-Marek, and e. a. Baringhaus, "Hot spots and transient pockets: predicting the determinants of small-molecule binding to a protein-protein interface.," *Journal of chemical information and modeling*, vol. 52, pp. 120–33, Jan. 2012.

[300] N. Bhardwaj, A. Abyzov, D. Clarke, C. Shou, and e. a. Gerstein, "Integration of protein motions with molecular networks reveals different mechanisms for permanent and transient interactions.," *Protein science : a publication of the Protein Society*, vol. 20, pp. 1745–54, Oct. 2011.

[301] A. Stein, M. Rueda, A. Panjkovich, M. Orozco, and e. a. Aloy, "A systematic study of the energetics involved in structural changes upon association and connectivity in protein interaction networks.," *Structure (London, England : 1993)*, vol. 19, pp. 881–9, June 2011.

[302] T. Meyer, M. D'Abramo, A. Hospital, M. Rueda, and e. a. Ferrer-Costa, "MoDEL (Molecular Dynamics Extended Library): a database of atomistic molecular dynamics trajectories.," *Structure (London, England : 1993)*, vol. 18, pp. 1399–409, Nov. 2010.

[303] T. Clackson and J. A. Wells, "A hot spot of binding energy in a hormone-receptor interface.," *Science (New York, N.Y.)*, vol. 267, pp. 383–6, Jan. 1995.

[304] B. S. Zerbe, D. R. Hall, S. Vajda, A. Whitty, and e. a. Kozakov, "Relationship between hot spot residues and ligand binding hot spots in protein-protein interfaces.," *Journal of chemical information and modeling*, vol. 52, pp. 2236–44, Aug. 2012.

[305] J. Fernández-Recio, "Prediction of protein binding sites and hot spots," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 1, pp. 680–698, Sept. 2011.

[306] D. R. Koes and C. J. Camacho, "PocketQuery: protein-protein interaction inhibitor starting points from protein-protein interaction structure.," *Nucleic acids research*, vol. 40, pp. W387–92, July 2012.

[307] D. Koes, K. Khoury, Y. Huang, W. Wang, and e. a. Bista, "Enabling large-scale design, synthesis and validation of small molecule protein-protein antagonists.," *PloS one*, vol. 7, p. e32839, Jan. 2012.

[308] B. O. Villoutreix, K. Bastard, O. Sperandio, R. Fahraeus, and e. a. Poyet, "In silico-in vitro screening of protein-protein interactions: towards the next generation of therapeutics.," *Current pharmaceutical biotechnology*, vol. 9, pp. 103–22, Apr. 2008.

[309] M. Floris and S. Moro, "Mimicking Peptides…In Silico," *Molecular Informatics*, vol. 31, pp. 12–20, Jan. 2012.

[310] P. M. Watt, "Screening for peptide drugs from the natural repertoire of biodiverse protein folds.," *Nature biotechnology*, vol. 24, pp. 177–83, Feb. 2006.

[311] P. Vanhee, A. M. van der Sloot, E. Verschueren, L. Serrano, and e. a. Rousseau, "Computational design of peptide ligands.," *Trends in biotechnology*, vol. 29, pp. 231–9, May 2011.

[312] P. Thiel, M. Kaiser, and C. Ottmann, "Small-molecule stabilization of protein-protein interactions: an underestimated concept in drug discovery?," *Angewandte Chemie (International ed. in English)*, vol. 51, pp. 2012–8, Feb. 2012.

[313] P. Block, N. Weskamp, A. Wolf, and G. Klebe, "Strategies to search and design stabilizers of protein-protein interactions: a feasibility study.," *Proteins*, vol. 68, pp. 170–86, July 2007.

[314] H. Jiang, K. E. Bower, A. E. Beuscher, B. Zhou, and e. a. Bobkov, "Stabilizers of the Max homodimer identified in virtual ligand screening inhibit Myc function.," *Molecular pharmacology*, vol. 76, pp. 491–502, Sept. 2009.

[315] S. S. Ray, R. J. Nowak, R. H. Brown, and P. T. Lansbury, "Small-molecule-mediated stabilization of familial amyotrophic lateral sclerosis-linked superoxide dismutase mutants against unfolding and aggregation.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, pp. 3639–44, Mar. 2005.

[316] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.," *Advanced drug delivery reviews*, vol. 46, pp. 3–26, Mar. 2001.

[317] O. Sperandio, C. H. Reynès, A.-C. Camproux, and B. O. Villoutreix, "Rationalizing the chemical space of protein-protein interaction inhibitors.," *Drug discovery today*, vol. 15, pp. 220–9, Mar. 2010.

[318] A. P. Higueruelo, A. Schreyer, G. R. J. Bickerton, W. R. Pitt, C. R. Groom, and T. L. Blundell, "Atomic interactions and profile of small molecules disrupting protein-protein interfaces: the TIMBAL database.," *Chemical biology & drug design*, vol. 74, pp. 457–67, Nov. 2009.

[319] F. Jiang, B. R. Frey, M. L. Evans, J. C. Friel, and e. a. Hopper, "Gene activation by dissociation of an inhibitor from a transcriptional activation domain.," *Molecular and cellular biology*, vol. 29, pp. 5604–10, Oct. 2009.

[320] M. R. Arkin and J. A. Wells, "Small-molecule inhibitors of protein-protein interactions: progressing towards the dream.," *Nature reviews. Drug discovery*, vol. 3, pp. 301–17, Apr. 2004.

[321] P. Weinkam, J. Pons, and A. Sali, "Structure-based model of allostery predicts coupling between distant sites.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, pp. 4875–80, Mar. 2012.

[322] J. Löwe, H. Li, K. H. Downing, and E. Nogales, "Refined structure of alpha beta-tubulin at 3.5 A resolution.," *Journal of molecular biology*, vol. 313, pp. 1045–57, Nov. 2001.

[323] P. J. Conn, A. Christopoulos, and C. W. Lindsley, "Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders.," *Nature reviews. Drug discovery*, vol. 8, pp. 41–54, Jan. 2009.

[324] P. M. Gasper, B. Fuglestad, E. A. Komives, P. R. L. Markwick, and e. a. McCammon, "Allosteric networks in thrombin distinguish procoagulant vs. anticoagulant activities.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, pp. 21216–22, Dec. 2012.

[325] P. Zhang, E. V. Smith-Nguyen, M. M. Keshwani, M. S. Deal, and e. a. Kornev, "Structure and allostery of the PKA RIIβ tetrameric holoenzyme.," *Science (New York, N.Y.)*, vol. 335, pp. 712–6, Feb. 2012.

[326] M. Good, G. Tang, J. Singleton, A. Reményi, and W. A. Lim, "The Ste5 scaffold directs mating signaling by catalytically unlocking the Fus3 MAP kinase for activation.," *Cell*, vol. 136, pp. 1085–97, Mar. 2009.

[327] A. D. Thompson, A. Dugan, J. E. Gestwicki, and A. K. Mapp, "Fine-tuning multiprotein complexes using small molecules.," *ACS chemical biology*, vol. 7, pp. 1311–20, Aug. 2012.

[328] Y. L. Deribe, T. Pawson, and I. Dikic, "Post-translational modifications in signal integration.," *Nature structural & molecular biology*, vol. 17, pp. 666–72, June 2010.

[329] R. Nussinov, C.-J. Tsai, F. Xin, and P. Radivojac, "Allosteric post-translational modification codes.," *Trends in biochemical sciences*, vol. 37, pp. 447–55, Oct. 2012.

[330] Y.-H. Ahn, Y. Hwang, H. Liu, X. J. Wang, and e. a. Zhang, "Electrophilic tuning of the chemoprotective natural product sulforaphane.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, pp. 9590–5, May 2010.

[331] B. Ma, M. Shatsky, H. J. Wolfson, and R. Nussinov, "Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations.," *Protein science : a publication of the Protein Society*, vol. 11, pp. 184–97, Feb. 2002.

[332] O. Hantschel, F. Grebien, and G. Superti-Furga, "The growing arsenal of ATP-competitive and allosteric inhibitors of BCR-ABL.," *Cancer research*, vol. 72, pp. 4890–5, Oct. 2012.

[333] G. R. Bowman and P. L. Geissler, "Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, pp. 11681–6, July 2012.

[334] Z. Huang, L. Zhu, Y. Cao, G. Wu, and e. a. Liu, "ASD: a comprehensive database of allosteric proteins and modulators.," *Nucleic acids research*, vol. 39, pp. D663–9, Jan. 2011.

[335] B. Ma and R. Nussinov, "Amplification of signaling via cellular allosteric relay and protein disorder.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, pp. 6887–8, Apr. 2009.

[336] M. A. Antal, C. Böde, and P. Csermely, "Perturbation waves in proteins and protein networks: applications of percolation and game theories in signaling and drug design.," *Current protein & peptide science*, vol. 10, pp. 161–72, Apr. 2009.

[337] R. Nussinov, C.-J. Tsai, and P. Csermely, "Allo-network drugs: harnessing allostery in cellular networks.," *Trends in pharmacological sciences*, vol. 32, pp. 686–93, Dec. 2011.

[338] G. M. Lee, T. Shahian, A. Baharuddin, J. E. Gable, and e. a. Craik, "Enzyme inhibition by allosteric capture of an inactive conformation.," *Journal of molecular biology*, vol. 411, pp. 999–1016, Sept. 2011.

[339] T. Ideker and N. J. Krogan, "Differential network biology.," *Molecular systems biology*, vol. 8, p. 565, Jan. 2012.

[340] R. A. Wilke and M. E. Dolan, "Genetics and variable drug response.," *JAMA*, vol. 306, pp. 306–7, July 2011.

[341] G. R. Abecasis, D. Altshuler, A. Auton, L. D. Brooks, R. M. Durbin, R. A. Gibbs, M. E. Hurles, and G. A. McVean, "A map of human genome variation from population-scale sequencing.," *Nature*, vol. 467, pp. 1061–73, Oct. 2010.

[342] K. R. Crews, J. K. Hicks, C.-H. Pui, M. V. Relling, and e. a. Evans, "Pharmacogenomics and individualized medicine: translating science into practice.," *Clinical pharmacology and therapeutics*, vol. 92, pp. 467–75, Oct. 2012.

[343] M. R. Nelson, D. Wegmann, M. G. Ehm, D. Kessner, P. St Jean, C. Verzilli, J. Shen, Z. Tang, S.-A. Bacanu, D. Fraser, L. Warren, J. Aponte, M. Zawistowski, X. Liu, H. Zhang, Y. Zhang, J. Li, Y. Li, L. Li, P. Woollard, S. Topp, M. D. Hall, K. Nangle, J. Wang, G. Abecasis, L. R. Cardon, S. Zöllner, J. C. Whittaker, S. L. Chissoe, J. Novembre, and V. Mooser, "An abundance

of rare functional variants in 202 drug target genes sequenced in 14,002 people.," *Science (New York, N.Y.)*, vol. 337, pp. 100–4, July 2012.

[344] E. C. Yiannakopoulou, "Pharmacogenomics of phase II metabolizing enzymes and drug transporters: clinical implications.," *The pharmacogenomics journal*, vol. 13, pp. 105–9, Apr. 2013.

[345] S. Preissner, K. Kroll, M. Dunkel, C. Senger, and e. a. Goldsobel, "SuperCYP: a comprehensive database on Cytochrome P450 enzymes including a tool for analysis of CYP-drug interactions.," *Nucleic acids research*, vol. 38, pp. D237–43, Jan. 2010.

[346] T.-D. Luu, A.-M. Rusu, V. Walter, R. Ripp, and e. a. Moulinier, "MSV3d: database of human MisSense Variants mapped to 3D protein structure.," *Database : the journal of biological databases and curation*, vol. 2012, p. bas018, Jan. 2012.

[347] J. L. Lahti, G. W. Tang, E. Capriotti, T. Liu, and e. a. Altman, "Bioinformatics and variability in drug response: a protein structural perspective.," *Journal of the Royal Society, Interface / the Royal Society*, vol. 9, pp. 1409–37, July 2012.

[348] T. Zhou, L. Parillon, F. Li, Y. Wang, and e. a. Keats, "Crystal structure of the T315I mutant of AbI kinase.," *Chemical biology & drug design*, vol. 70, pp. 171–81, Sept. 2007.

[349] W. W. Chan, S. C. Wise, M. D. Kaufman, Y. M. Ahn, and e. a. Ensinger, "Conformational control inhibition of the BCR-ABL1 tyrosine kinase, including the gatekeeper T315I mutant, by the switch-control inhibitor DCC-2036.," *Cancer cell*, vol. 19, pp. 556–68, Apr. 2011.

[350] E. Sano, W. Li, H. Yuki, X. Liu, and e. a. Furihata, "Mechanism of the decrease in catalytic activity of human cytochrome P450 2C9 polymorphic variants investigated by computational analysis.," *Journal of computational chemistry*, vol. 31, pp. 2746–58, Nov. 2010.

[351] G.-M. Tan, E. Wu, Y.-Y. Lam, and B. P. Yan, "Role of warfarin pharmacogenetic testing in clinical practice.," *Pharmacogenomics*, vol. 11, pp. 439–48, Mar. 2010.

[352] C. Kiel and L. Serrano, "Cell type-specific importance of ras-c-raf complex association rate constants for MAPK signaling.," *Science signaling*, vol. 2, p. ra38, Jan. 2009.

[353] T. M. K. Cheng, L. Goehring, L. Jeffery, Y.-E. Lu, and e. a. Hayles, "A structural systems biology approach for quantifying the systemic consequences of missense mutations in proteins.," *PLoS computational biology*, vol. 8, p. e1002738, Jan. 2012.

[354] K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, and e. a. Wolf-Watz, "Intrinsic motions along an enzymatic reaction trajectory.," *Nature*, vol. 450, pp. 838–44, Dec. 2007.

[355] E. Z. Eisenmesser, D. A. Bosco, M. Akke, and D. Kern, "Enzyme dynamics during catalysis.," *Science (New York, N.Y.)*, vol. 295, pp. 1520–3, Feb. 2002.

[356] C. Micheletti, "Comparing proteins by their internal dynamics: exploring structure-function relationships beyond static structural alignments.," *Physics of life reviews*, vol. 10, pp. 1–26, Mar. 2013.

[357] J. J. Falke, "Enzymology. A moving story.," *Science (New York, N.Y.)*, vol. 295, pp. 1480–1, Feb. 2002.

[358] M. Orozco, "A theoretical view of protein dynamics.," 2014.

[359] J. A. Velázquez-Muriel, M. Rueda, I. Cuesta, A. Pascual-Montano, M. Orozco, and J.-M. Carazo, "Comparison of molecular dynamics and superfamily spaces of protein domain deformation.," *BMC structural biology*, vol. 9, p. 6, Jan. 2009.

[360] A. Leo-Macias, P. Lopez-Romero, D. Lupyan, D. Zerbino, and e. a. Ortiz, "An analysis of core deformations in protein superfamilies.," *Biophysical journal*, vol. 88, pp. 1291–9, Feb. 2005.

[361] I. Bahar, T. R. Lezon, A. Bakan, and I. H. Shrivastava, "Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins.," *Chemical reviews*, vol. 110, pp. 1463–97, Mar. 2010.

[362] A. Das, M. Gur, M. H. Cheng, S. Jo, and e. a. Bahar, "Exploring the conformational transitions of biomolecular systems using a simple two-state anisotropic network model.," *PLoS computational biology*, vol. 10, p. e1003521, Apr. 2014.

[363] J. A. McCammon, B. R. Gelin, and M. Karplus, "Dynamics of folded proteins.," *Nature*, vol. 267, pp. 585–90, June 1977.

[364] M. K. Kim, R. L. Jernigan, and G. S. Chirikjian, "Efficient generation of feasible pathways for protein conformational transitions.," *Biophysical journal*, vol. 83, pp. 1620–30, Sept. 2002.

[365] L. Yang, G. Song, and R. L. Jernigan, "How well can we understand large-scale protein motions using normal modes of elastic network models?," *Biophysical journal*, vol. 93, pp. 920–9, Aug. 2007.

[366] R. O. Dror, R. M. Dirks, J. P. Grossman, H. Xu, and e. a. Shaw, "Biomolecular simulation: a computational microscope for molecular biology.," *Annual review of biophysics*, vol. 41, pp. 429–52, Jan. 2012.

[367] T. Shimamura, S. Weyand, O. Beckstein, N. G. Rutherford, and e. a. Hadden, "Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1.," *Science (New York, N.Y.)*, vol. 328, pp. 470–3, Apr. 2010.

[368] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, and e. a. Dror, "Atomic-level characterization of the structural dynamics of proteins.," *Science (New York, N.Y.)*, vol. 330, pp. 341–6, Oct. 2010.

[369] R. Elber, "A milestoning study of the kinetics of an allosteric transition: atomically detailed simulations of deoxy Scapharca hemoglobin.," *Biophysical journal*, vol. 92, pp. L85–7, May 2007.

[370] R. Elber, "Long-timescale simulation methods.," *Current opinion in structural biology*, vol. 15, pp. 151–6, Apr. 2005.

[371] R. Elber and A. West, "Atomically detailed simulation of the recovery stroke in myosin by Milestoning.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, pp. 5001–5, Mar. 2010.

[372] J. R. Perilla, O. Beckstein, E. J. Denning, and T. B. Woolf, "Computing ensembles of transitions from stable states: Dynamic importance sampling.," *Journal of computational chemistry*, vol. 32, pp. 196–209, Jan. 2011.

[373] A. Laio and M. Parrinello, "Escaping free-energy minima.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 12562–6, Oct. 2002.

[374] P. Sfriso, A. Emperador, L. Orellana, A. Hospital, J. L. Gelpí, and M. Orozco, "Finding Conformational Transition Pathways from Discrete Molecular Dynamics Simulations," *Journal of Chemical Theory and Computation*, vol. 8, pp. 4707–4718, Nov. 2012.

[375] P.-c. Chen and J. S. Hub, "Validating solution ensembles from molecular dynamics simulation by wide-angle X-ray scattering data.," *Biophysical journal*, vol. 107, pp. 435–47, July 2014.

[376] D. Seeliger and B. L. de Groot, "Conformational transitions upon ligand binding: holo-structure prediction from apo conformations.," *PLoS computational biology*, vol. 6, p. e1000634, Jan. 2010.

[377] H. van den Bedem, G. Bhabha, K. Yang, P. E. Wright, and e. a. Fraser, "Automated identification of functional dynamic contact networks from X-ray crystallography," *Nature Methods*, vol. 10, pp. 896–902, Aug. 2013.

[378] O. Beckstein, E. J. Denning, J. R. Perilla, and T. B. Woolf, "Zipping and unzipping of adenylate kinase: atomistic insights into the ensemble of open<–>closed transitions.," *Journal of molecular biology*, vol. 394, pp. 160–76, Nov. 2009.

[379] P. C. Whitford, O. Miyashita, Y. Levy, and J. N. Onuchic, "Conformational transitions of adenylate kinase: switching by cracking.," *Journal of molecular biology*, vol. 366, pp. 1661–71, Mar. 2007.

[380] D. R. Weiss and M. Levitt, "Can morphing methods predict intermediate structures?," *Journal of molecular biology*, vol. 385, pp. 665–74, Jan. 2009.

[381] D. T. Jones, T. Singh, T. Kosciolek, and S. Tetchner, "MetaPSICOV: combining coevolution methods for accurate prediction of contacts and long range hydrogen bonding in proteins.," *Bioinformatics (Oxford, England)*, vol. 31, pp. 999–1006, Apr. 2015.

[382] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, and e. a. Marks, "Direct-coupling analysis of residue coevolution captures native contacts across many protein families.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, pp. E1293–301, Dec. 2011.

[383] M. Michel, S. Hayat, M. J. Skwark, C. Sander, and e. a. Marks, "Pcons-Fold: improved contact predictions improve protein models.," *Bioinformatics (Oxford, England)*, vol. 30, pp. i482–8, Sept. 2014.

[384] T. A. Hopf, S. Morinaga, S. Ihara, K. Touhara, and e. a. Marks, "Amino acid coevolution reveals three-dimensional structure and functional domains of insect odorant receptors.," *Nature communications*, vol. 6, p. 6077, Jan. 2015.

[385] F. Morcos, B. Jana, T. Hwa, and J. N. Onuchic, "Coevolutionary signals across protein lineages help capture multiple protein conformations.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. 20533–8, Dec. 2013.

[386] M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, and e. a. Hwa, "Identification of direct residue contacts in protein-protein interaction by message passing," *Proceedings of the National Academy of Sciences*, vol. 106, pp. 67–72, Dec. 2008.

[387] E. A. Proctor, F. Ding, and N. V. Dokholyan, "Discrete molecular dynamics - WIREs Computational Molecular Science."

[388] P. Sfriso, A. Emperador, J. L. Gelpí, and M. Orozco, "Computational Approaches to Protein Dynamics: From Quantum to Coarse-Grained Methods - CRC Press Book."

[389] H. Taketomi, F. Kanô, and N. Go, "The effect of amino acid substitution on protein-folding and -unfolding transition studied by computer simulation.," *Biopolymers*, vol. 27, pp. 527–59, Apr. 1988.

[390] Y. Ueda, H. Taketomi, and N. G?, "Studies on protein folding, unfolding, and fluctuations by computer simulation. II. A. Three-dimensional lattice model of lysozyme," *Biopolymers*, vol. 17, pp. 1531–1548, June 1978.

[391] V. Tozzini, "Coarse-grained models for proteins.," *Current opinion in structural biology*, vol. 15, pp. 144–50, Apr. 2005.

[392] Y. Hisanaga, H. Ago, N. Nakagawa, K. Hamada, and e. a. Ida, "Structural basis of the substrate-specific two-step catalysis of long chain fatty acyl-CoA synthetase dimer.," *The Journal of biological chemistry*, vol. 279, pp. 31717–26, July 2004.

[393] R. D. Finn, A. Bateman, J. Clements, P. Coggill, and e. a. Eberhardt, "Pfam: the protein families database.," *Nucleic acids research*, vol. 42, pp. D222–30, Jan. 2014.

[394] K. Khafizov, C. Madrid-Aliste, S. C. Almo, and A. Fiser, "Trends in structural coverage of the protein universe and the impact of the Protein Structure Initiative.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, pp. 3733–8, Mar. 2014.

[395] H. Kamisetty, S. Ovchinnikov, and D. Baker, "Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, pp. 15674–9, Sept. 2013.

[396] D. de Juan, F. Pazos, and A. Valencia, "Emerging methods in protein co-evolution," *Nature Reviews Genetics*, vol. 14, pp. 249–261, Mar. 2013.

[397] M. Remmert, A. Biegert, A. Hauser, and J. Söding, "HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment.," *Nature methods*, vol. 9, pp. 173–5, Feb. 2012.

[398] A. Emperador, O. Carrillo, M. Rueda, and M. Orozco, "Exploring the suitability of coarse-grained techniques for the representation of protein dynamics.," *Biophysical journal*, vol. 95, pp. 2127–38, Sept. 2008.

[399] D. Shirvanyants, F. Ding, D. Tsao, S. Ramachandran, and e. a. Dokholyan, "Discrete molecular dynamics: an efficient and versatile simulation method for fine protein characterization.," *The journal of physical chemistry. B*, vol. 116, pp. 8375–82, July 2012.

[400] M. Orozco, L. Orellana, A. Hospital, A. N. Naganathan, and e. a. Emperador, "Coarse-grained representation of protein flexibility. Foundations, successes, and shortcomings.," *Advances in protein chemistry and structural biology*, vol. 85, pp. 183–215, Jan. 2011.

[401] P. Sfriso, A. Hospital, A. Emperador, and M. Orozco, "Exploration of conformational transition pathways from coarse-grained simulations.," *Bioinformatics (Oxford, England)*, vol. 29, pp. 1980–6, Aug. 2013.

[402] A. Barducci, M. Bonomi, and M. Parrinello, "Metadynamics," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 1, pp. 826–843, Sept. 2011.

[403] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," *KDD-96 Proceedings*, 1996.

[404] X. Daura, K. Gademann, B. Jaun, D. Seebach, and e. a. van Gunsteren, "Peptide Folding: When Simulation Meets Experiment," *Angewandte Chemie International Edition*, vol. 38, pp. 236–240, Jan. 1999.

[405] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation," *Journal of Chemical Theory and Computation*, vol. 4, pp. 435–447, Mar. 2008.

[406] J. Camps, O. Carrillo, A. Emperador, L. Orellana, A. Hospital, M. Rueda, D. Cicin-Sain, M. D'Abramo, J. L. Gelpí, and M. Orozco, "FlexServ: an integrated tool for the analysis of protein flexibility.," *Bioinformatics (Oxford, England)*, vol. 25, pp. 1709–10, July 2009.

[407] L. Orellana, M. Rueda, C. Ferrer-Costa, J. R. Lopez-Blanco, and e. a. Chacon, "Approaching Elastic Network Models to Molecular Dynamics Flexibility," *Journal of Chemical Theory and Computation*, vol. 6, pp. 2910–2923, Sept. 2010.

[408] C. Clementi, H. Nymeyer, and J. N. Onuchic, "Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins.," *Journal of molecular biology*, vol. 298, pp. 937–53, May 2000.

[409] A. L. Hopkins, J. S. Mason, and J. P. Overington, "Can we rationally design promiscuous drugs?," *Current opinion in structural biology*, vol. 16, pp. 127–36, Feb. 2006.

[410] M. L. Bolognesi, "Polypharmacology in a single drug: multitarget drugs.," *Current medicinal chemistry*, vol. 20, pp. 1639–45, Jan. 2013.

[411] X. Jalencas and J. Mestres, "Identification of Similar Binding Sites to Detect Distant Polypharmacology," *Molecular Informatics*, vol. 32, pp. 976–990, Dec. 2013.

[412] X. Lin, X.-P. Huang, G. Chen, R. Whaley, and e. a. Peng, "Life Beyond Kinases: Structure-Based Discovery of Sorafenib as Nanomolar Antagonist of 5-HT Receptors," *Journal of Medicinal Chemistry*, vol. 55, pp. 5749–5759, June 2012.

[413] A. A. Antolín, X. Jalencas, J. Yélamos, and J. Mestres, "Identification of Pim Kinases as Novel Targets for PJ34 with Confounding Effects in PARP Biology," *ACS Chemical Biology*, vol. 7, pp. 1962–1967, Dec. 2012.

[414] F. Prati, A. De Simone, P. Bisignano, A. Armirotti, M. Summa, D. Pizzirani, R. Scarpelli, D. I. Perez, V. Andrisano, A. Perez-Castillo, B. Monti, F. Massenzio, L. Polito, M. Racchi, A. D. Favia, G. Bottegoni, A. Martinez, M. L. Bolognesi, and A. Cavalli, "Multitarget drug discovery for Alzheimer's disease: triazinones as BACE-1 and GSK-3β inhibitors.," *Angewandte Chemie (International ed. in English)*, vol. 54, pp. 1578–82, Jan. 2015.

[415] S. J. Lee and J. Y. J. Wang, "Exploiting the promiscuity of imatinib.," *Journal of biology*, vol. 8, p. 30, Jan. 2009.

[416] H. Zhou, M. Gao, and J. Skolnick, "Comprehensive prediction of drug-protein interactions and side effects for the human proteome.," *Scientific reports*, vol. 5, p. 11090, Jan. 2015.

[417] L. Siragusa, S. Cross, M. Baroni, L. Goracci, and e. a. Cruciani, "BioGPS: navigating biological space to predict polypharmacology, off-targeting, and selectivity.," *Proteins*, vol. 83, pp. 517–32, Mar. 2015.

[418] J. Konc and D. Janežič, "Binding site comparison for function prediction and pharmaceutical discovery.," *Current opinion in structural biology*, vol. 25, pp. 34–9, Apr. 2014.

[419] L. Xie, L. Xie, and P. E. Bourne, "Structure-based systems biology for analyzing off-target binding.," *Current opinion in structural biology*, vol. 21, pp. 189–99, Apr. 2011.

[420] M. T. Wong, S. B. Choi, C. S. Kuan, S. L. Chua, and e. a. Chang, "Structural modeling and biochemical characterization of recombinant KPN_02809, a zinc-dependent metalloprotease from Klebsiella pneumoniae MGH 78578.," *International journal of molecular sciences*, vol. 13, pp. 901–17, Jan. 2012.

[421] C. Rubio-Perez, D. Tamborero, M. Schroeder, A. Antolín, J. Deu-Pons, C. Perez-Llamas, J. Mestres, A. Gonzalez-Perez, and N. Lopez-Bigas, "In Silico Prescription of Anticancer Drugs to Cohorts of 28 Tumor Types Reveals Targeting Opportunities," *Cancer Cell*, vol. 27, pp. 382–396, Mar. 2015.

[422] J. A. Capra, R. A. Laskowski, J. M. Thornton, M. Singh, and e. a. Funkhouser, "Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure.," *PLoS computational biology*, vol. 5, p. e1000585, Dec. 2009.

[423] R. I. Brinkworth, R. A. Breinl, and B. Kobe, "Structural basis and prediction of substrate specificity in protein serine/threonine kinases," *Proceedings of the National Academy of Sciences*, vol. 100, pp. 74–79, Dec. 2002.

[424] J. A. DiMasi and L. B. Faden, "Competitiveness in follow-on drug R&D: a race or imitation?," *Nature reviews. Drug discovery*, vol. 10, pp. 23–7, Jan. 2011.

[425] A. J. Venkatakrishnan, X. Deupi, G. Lebon, C. G. Tate, and e. a. Schertler, "Molecular signatures of G-protein-coupled receptors.," *Nature*, vol. 494, pp. 185–94, Feb. 2013.

[426] J. Jia, F. Zhu, X. Ma, Z. Cao, and e. a. Cao, "Mechanisms of drug combinations: interaction and network perspectives.," *Nature reviews. Drug discovery*, vol. 8, pp. 111–28, Feb. 2009.

[427] J. Gu, X. Zhang, Y. Ma, N. Li, and e. a. Luo, "Quantitative modeling of dose–response and drug combination based on pathway network," *Journal of Cheminformatics*, vol. 7, p. 19, May 2015.

[428] Y. Liu, Q. Wei, G. Yu, W. Gai, and e. a. Li, "DCDB 2.0: a major update of the drug combination database.," *Database : the journal of biological databases and curation*, vol. 2014, p. bau124, Jan. 2014.

[429] E. Becker, B. Robisson, C. E. Chapple, A. Guénoche, and C. Brun, "Multifunctional proteins revealed by overlapping clustering in protein interaction network.," *Bioinformatics (Oxford, England)*, vol. 28, pp. 84–90, Jan. 2012.

[430] T. Korcsmáros, M. S. Szalay, C. Böde, I. A. Kovács, and P. Csermely, "How to design multi-target drugs.," *Expert opinion on drug discovery*, vol. 2, pp. 799–808, June 2007.

[431] P. Csermely, V. Agoston, and S. Pongor, "The efficiency of multi-target drugs: the network approach might help drug design.," *Trends in pharmacological sciences*, vol. 26, pp. 178–82, Apr. 2005.

[432] S. D. Ghiassian, J. Menche, and A.-L. Barabási, "A DIseAse MOdule Detection (DIAMOnD) Algorithm Derived from a Systematic Analysis of Connectivity Patterns of Disease Proteins in the Human Interactome," *PLOS Computational Biology*, vol. 11, p. e1004120, Apr. 2015.

[433] F. Azuaje, Y. Devaux, and D. R. Wagner, "Identification of potential targets in biological signalling systems through network perturbation analysis.," *Bio Systems*, vol. 100, pp. 55–64, Apr. 2010.

[434] G. Yadav and S. Babu, "NEXCADE: perturbation analysis for complex networks.," *PloS one*, vol. 7, p. e41827, Jan. 2012.

[435] M. D. M. Leiserson, F. Vandin, H.-T. Wu, J. R. Dobson, and e. a. Eldridge, "Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes," *Nature Genetics*, vol. 47, pp. 106–14, Dec. 2014.

[436] R. A. Cairns, I. S. Harris, and T. W. Mak, "Regulation of cancer cell metabolism.," *Nature reviews. Cancer*, vol. 11, pp. 85–95, Feb. 2011.

[437] V. J. Haupt and M. Schroeder, "Old friends in new guise: repositioning of known drugs with structural bioinformatics.," *Briefings in bioinformatics*, vol. 12, pp. 312–26, July 2011.

[438] R. Mosca, T. Pons, A. Céol, A. Valencia, and P. Aloy, "Towards a detailed atlas of protein-protein interactions.," *Current opinion in structural biology*, vol. 23, pp. 929–40, Dec. 2013.

[439] L. S. Johnson, S. R. Eddy, and E. Portugaly, "Hidden Markov model speed heuristic and iterative HMM search procedure.," *BMC bioinformatics*, vol. 11, p. 431, Jan. 2010.

[440] H. Cheng, R. D. Schaeffer, Y. Liao, L. N. Kinch, and e. a. Pei, "ECOD: an evolutionary classification of protein domains.," *PLoS computational biology*, vol. 10, p. e1003926, Dec. 2014.

[441] A. P. Burgard, P. Pharkya, and C. D. Maranas, "Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization," *Biotechnology and Bioengineering*, vol. 84, pp. 647–657, Dec. 2003.

[442] L. Xie, X. Ge, H. Tan, L. Xie, and e. a. Zhang, "Towards structural systems pharmacology to study complex diseases and personalized medicine.," *PLoS computational biology*, vol. 10, p. e1003554, May 2014.

[443] N. J. Schork, "Personalized medicine: Time for one-person trials," *Nature*, vol. 520, pp. 609–611, Apr. 2015.

[444] T. Juan-Blanco, M. Duran-Frigola, and P. Aloy, "IntSide: a web server for the chemical and biological examination of drug side effects.," *Bioinformatics (Oxford, England)*, vol. 31, pp. 612–3, Feb. 2015.

[445] R. Arroyo, M. Duran-Frigola, C. Berenguer, M. Soler-López, and P. Aloy, "Charting the molecular links between driver and susceptibility genes in colorectal cancer.," 2014.

[446] R. Arroyo, G. Suñé, A. Zanzoni, M. Duran-Frigola, V. Alcalde, T. H. Stracker, M. Soler-López, and P. Aloy, "Systematic identification of molecular links between core and candidate genes in breast cancer.," 2015.