Drug design at biological systems level

Nikita Remez Vinogradov

Doctoral Thesis UPF / 2015

Thesis director:

Dr. Jordi Mestres

CEXS Department

Part of the research in this Thesis has been carried out at the Systems Pharmacology Research Group, within the Research Programme on Biomedical Informatics (GRIB) at Parc de Recerca Biomèdica de Barcelona (PRBB).

Part of the research in this Thesis has been carried out at Chemotargets.

The research presented in this Thesis has been supported by the Catalan Government grant AGAUR TEM-DGR 2010.

Acknowledgements / Agraïments / Благодарности

En primer lloc, m'agradaria donar les gràcies a la persona que em va donar l'oportunitat de treballar al camp que realment m'encanta, em va guiar i em va donar suport durant tot el procés – Doctor Jordi Mestres. A veure si és veritat que venim de la mateixa tribu?

A tots els companys i companyes que m'havien acompanyat durant aquest viatge científic – des dels primers dies feliços a Biotec – "als de la Vila!" (Alberto, Gemma, Jurdan, Marta, Maria, Natalia i Sergio); dels 2 ràpids anys al Màster, especialment Marc, Michi i Ruth. A tots els que m'havien acollit al GRIB, en ordre d'aparició: Carina, Ferran (a veure si quedem per fer més partides!), Albert, Muyayo Ingo, Xavi, Praveena, Montse, Maricarmen, Ángel, Cristian... crec que em deixo unes 10 persones més, però aquesta part la escric al final de tot i el meu cervell ja no dóna més de si. Ricard, no apareixes en aquesta llista per que et mereixes una menció especial, per suport moral i científic que m'has donat des dels primers dies que vaig aparèixer al *zulo de los recienllegados*, fins els últims moments de la escriptura d'aquest manuscrit. També estic enormement agraït al veterà de Chemotargets, David, pels valuosos consells sobre química computacional i sobre la realitat empresarial.

Des de que vaig acabar el batxillerat, sempre havia volgut agrair a la meva professora de biologia, Virgínia, per ensenyar-me el pensament crític i demostrar que la ciència és impressionant.

Now, I will continue in English, so that everybody could understand how important is this person for me. Katia, my perfect wife and dear soulmate (too sweet? but that's true!), you've made so much for me

during all these years, that, in case I succeed, I will have to share the PhD title with you. Твоя забота, любовь и поддержка так сильно значат для меня, что невозможно описать ни на каком языке, и сейчас я не уверен правильно ли я написал “ни на каком”, но главное, что я тебя люблю. I LOVE YOU. T’estimo!

Огромное спасибо моим вторым родителям, Надежде и Владимиру. Во-первых, за свою прекрасную дочь, что делает меня счастливым каждый день, а также за вашу доброту, чуткость и внимание.

Большое спасибо, конечно, всей моей огромной семье. Всем бабушкам и дедушкам – вы мои корни, я всегда о вас помню и росту из вас. Всем-всем двоюродным, троюродным и особенно четвероюродным братьям и сёстрам, благодарю вас за то, что верили в меня и подбадирали. Конечно, спасибо Мите и Лене за то, что всегда были рядом.

Мама и папа, думаете я про вас забыл? Вот и нет, просто оставил на десерт, как самых важных людей в моей жизни. Благодаря вам, я сейчас такой какой есть, спасибо вам за... да всего и не перечислить, главное за вашу любовь и душевную поддержку. А папе – отдельное спасибо за тягу к науке!

Table of Contents

Abstract.....xi

Resum.....xii

Preface.....xiii

Part I: Introduction.....1

I.1 Drug discovery issues.....3

I.2 New strategies in drug discovery.....6

I.3 Target-centered systems biology approach.....17

I.4 Particularities of PhD research in a computational biotech company.....29

I.5 References.....31

Part II: Objectives.....43

Part III: Results.....47

III.1 A Chemocentric Approach to the Identification of Cancer Targets.....50

III.2 Design of a General-Purpose European Compound Screening Library for EU-OPENSCREEN.....86

III.3 Computational prediction of off-target related safety liabilities of molecules: Cardiotoxicity, hepatotoxicity and reproductive toxicity.....150

III.4 Large-Scale Predictive Drug Safety: From Structural Alerts to Biological Mechanisms.....156

III.5 The *In Vitro* Pharmacological Profile of Drugs as a Proxy Indicator of Potential *In Vivo* Organ Toxicities.....203

III.6 CT-link distribution package report.....251

Part IV: Discussions.....259

Part V: Conclusions.....265

Appendix.....271

Abstract

The explosion of reductionist approaches at the end of the XXth century allowed for fast and high-throughput data collection in pharmaceutical industry, but did not deliver the expected gain in drug discovery performance. Omics methodologies were able to provide large amounts of information on one-to-one cause-effect relationships, but could not explain some of the complex effects in living organisms and biological systems in general. This Thesis, performed in the premises of an emerging biotech company, represents an attempt to explore the impact of an integrative systems approach to drug discovery. On the one hand, *in silico* prediction of drug-target interactions developed at Chemotargets was applied to the identification of cancer-relevant targets, the design of a biologically diverse chemical library, and the implementation of a novel methodology for predictive drug safety. On the other hand, major efforts were devoted to develop and carefully validate a new approach to anticipating *in vivo* organ toxicities from *in vitro* pharmacological profiles and gene expression datas. This novel methodology was designed to complement and assist the expert toxicologist by providing insights at the anatomical level . Last but no least, key contributions have been made to develop a brand new platform-independent version of the company's flagship software, CT-link, allowing for easy distribution and commercialization.

Resum

A les darreries del segle XX, l'auge de les aproximacions reduccionistes van permetre a la indústria farmacèutica la recopilació de gran quantitat de informació, però l'impacte en el rendiment de la producció de nous fàrmacs no va ser l'esperat. Encara que es van extreure moltes dades sobre les relacions un-a-un entre entitats sistèmiques els efectes complexos causats als sistemes biològics no es van poder adreçar adequadament. Aquesta Tesi, desenvolupada en el marc d'una empresa biotecnològica emergent, intenta introduir un marc de referència integral per l'aproximació sistèmica al disseny de fàrmacs. D'una banda, s'ha aplicat la predicció *in silico* de xarxes fàrmac-diana per identificar les proteïnes relacionades amb càncer, per la construcció d'una biblioteca química d'àmplia cobertura biològica i per anticipar la toxicitat relacionada amb dianes secundàries. D'altra banda, es va ampliar l'aproximació de biologia de sistemes per abastar les connexions anatòmiques. Aquesta nova metodologia fou dissenyada per complementar i assistir al toxicòleg expert en la identificació de toxicitats a nivell anatòmic. Finalment, durant aquests anys s'ha contribuït de manera important al desenvolupament d'una nova versió independent de la plataforma del programari estrella de l'empresa, CT-link, que n'ha facilitat enormement la seva distribució i comercialització.

Preface

Approaches to drug design have had various breakthroughs in history. Pharmaceutical industry evolved in parallel with and received impulse from scientific advances in different areas. Progress in organic chemistry allowed to jump from search of active ingredients within natural extracts to synthetic compounds; progress in biology and zoology allowed to incorporate laboratory animal tests in efficacy and toxicity studies; progress in genomics, proteomics and biochemistry provided reductionist methodologies for development of selective drugs. Now we are experiencing yet another breakthrough, actually two simultaneously. Advances in integrative approaches, such as systems biology and chemogenomics, may lead to more rational drug design and discovery of more effective and safer treatments. Additionally, exponential growth of data obtained from high-throughput experiments, together with progress in computer technologies, provide new in silico tools that are able to analyse and model complex relationships that are needed to be addressed for studies of the aforementioned biological systems.

The opportunity to start this Thesis in the Systems Pharmacology Group and to continue its development in the computationally-oriented biotech company Chemotargets allowed me to take part in one of the current synergetic breakthroughs. The main aim of the Thesis was to design a new systems approach to drug design and to contribute to its development and application to large-scale predictive drug safety. To achieve this objective, this Thesis has been

divided in five parts. In the first introductory part, the current state of drug discovery is reviewed, considering its problems and its advances. The corresponding references are located afterwards. In the second part, the specific objectives of the research undertaken during my PhD are listed. Next, the results of the work are presented in the form of scientific papers, most of them published already, and a description of the CT-link software distribution version is presented. Finally, discussion of the research results and conclusions derived from them can be found in the last two parts of the manuscript.

Part I: Introduction

I.1 Drug discovery issues

Drug discovery is a rather conservative process. Despite of a gradual, but slow, implementation of innovative experimental techniques, adoption of new computational methods occurs very rarely. The advent of high-throughput methodologies biased the process towards gathering large volumes of data rather than favoring a culture of understanding the human body as a complex biological system.

Recent advances in drug discovery have not reduced the total cost of bringing new molecular entities to the market [1]. Since 2003, the number of molecules in development has increased by 62% and research expenditure has doubled, but the total number of approved molecules has decreased by 35% compared to the last decade. A recent study among 835 pharma companies (covering the whole spectrum from emerging biotech to large pharma) demonstrated that only 10.4% from a total of 7372 indication development paths in Phase I got finally FDA approval [2]. If only indications related to new small-molecule entities were considered, the success rate would be even smaller (7.6% from a total of 1335). Most surprisingly, indications in development that are related to “major” diseases (oncological, cardiovascular, neurological and respiratory) had less success rate than the others.

Internal studies carried out by AstraZeneca corroborate the reduction of R&D productivity [3]. The summary of the results from these studies is depicted in Figure 1. The investigation detected that 94 out of a total of 142 projects evaluated during 2005-2010 failed at the testing stage. More than a half of failures were due to safety

concerns. In the preclinical phase, 75% of safety issues were related to "off-target" activity. In later phases, the involvement of the primary target in drug safety increased. Another reason of failure in clinical phases was the lack of efficacy. AstraZeneca researchers attribute this to the lack of scientific knowledge on the link between the target and disease.

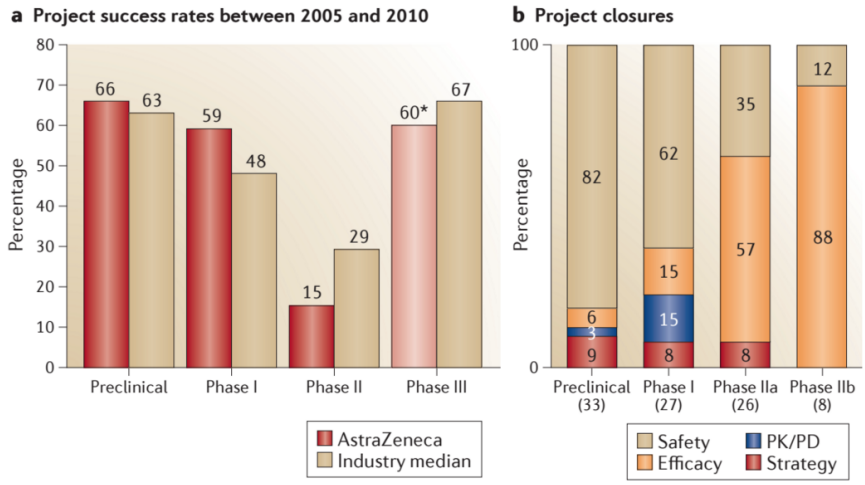

Figure 1. From [3]
Overview of project success rates and reasons for closure.
a) Overall project success rates for the AstraZeneca portfolio during the 2005–2010 period compared to the Pharmaceutical Benchmarking Forum (PBF) data. b) Stacked column plot showing primary reasons for project closure. Project closures were classified as failing because of safety (toxicology or clinical safety), efficacy (failure to achieve sufficient efficacy), pharmacokinetics/pharmacodynamics (PK/PD) or the strategy.

The commitment to the development of "me-too" drugs, as a source of short-term revenue, finally translates to a lack of innovative approaches in creation of new pharmaceuticals [4]. But, in contrast, investment on developing novel drugs can bring high value to the companies in the long term. Such a minor increase in FDA approvals during the last 10-15 years may be explained by the introduction of

advanced methodologies in the same time frame. Hopefully, the true impact of these new methodologies in productivity will be appreciated in the next decade.

All previous studies done on the matter [3,4] agree that drug discovery and development must rely on a wider knowledge of human biology and better understanding of the involvement of drug targets in diseases. But this approach has its own difficulties. The traditional paradigm of one gene — one disease has been overcome and, in order to identify the mechanistic target, one needs to build well-described molecular models for a disease state, and to demonstrate that affecting the target proposed will have a true therapeutical effect in humans [5]. Biological arguments based on association and correlation expression studies can help in target identification, but are weak to support target validation. There are several genetical techniques that help setting up similar-to-disease conditions in model organisms (siRNA, RNAi, conditional knock-outs etc..) and indicate which targets may have therapeutic potential. Because of the limited information available on the biological functions of druggable targets and their variability across tissues, the high-throughput target-based approach should also be combined with lower-throughput physiology-based approaches in the proof-of-principle stages.

I.2 New strategies in drug discovery

I.2.1 Computational drug discovery

In the last decades, computational approaches have been gradually added to the arsenal of technologies along the drug discovery workflow. *In silico* generated information is usually based on statistical analyses and processing of known data. Predictions and simulations are not widely accepted as decision makers, but are used as guidance for assisting prioritization.

Generally speaking, Computer-Assisted Drug Design (CADD) is divided in two classes: 1) structure-based and 2) ligand-based. The first type, as its name suggests, relies on the information that can be obtained from the structure of the target protein and can be used in the calculation of interaction energies between small molecules and the target active site. Meanwhile, the second type makes use of the information related to ligands (both active and inactive ones) and performs chemical similarity comparisons and construction quantitative structure-activity relation (QSAR) models [6]. The LB-CADD (ligand-based CADD) relies on the structural features of the reference compounds that are known to have activity on desired targets and analyses them in order to extract knowledge about activity. As a consequence, current LB-CADD methods are considered indirect approaches because they do not make use of the structural information on the target protein. LB-CADD basis is the so-called Similar-Property Principle formulated by Johnson and Maggiora in 1992 [7]. The principle states that structurally similar molecules are prone to have similar properties.

The complexity of some of the efforts to determine protein structures for certain families of high therapeutic interest (such as G protein-coupled receptors), by X-ray crystallography and NMR spectroscopy [8], represents a bottleneck in the applicability and coverage of SB-CADD methods in drug design. In this respect, advances in comparative modelling of protein structures have filled the gap in many instances [9–12]. Furthermore, active compounds obtained from ligand-based virtual high-throughput screenings were found to be more potent than those obtained from structure-based screenings [13,14].

LB-CADD methods rely on various approaches to describe the presumed essential chemical features of bioactive small molecules using *in silico* algorithms that often have to be optimized for computational cost and quality of output. The most appropriate feature-descriptors of choice will very much depend, on the one hand, on the therapeutical function predicted and, on the other hand, on the type of the ligand-based approach employed. Consequently, a wide range of algorithms for extracting chemical information have been developed and applied [15,16].

Molecular descriptors can be divided by their nature in either structural or physicochemical. They can be also classified according to the level of complexity:

- **0D**, single-value descriptors associated with the complete molecular structure. For example: molecular mass, number of atoms, log*P* and topological polar surface area (TPSA)

- **1D,** based on local connectivity, for example count of cycles or fragments of a determined length
- **2D**, numerical properties that can be calculated from the connection table representation of a molecule. The are also called “graph invariants”
- **2.5D**, combination of topological 2D descriptors and information from 3D structures, without considering the positions of atoms in 3D space explicitly
- **3D,** depend on the three-dimensional structure of the molecule and vary with conformational changes

Quantitative structure-activity relationship (QSAR) models are numerical regression methods that try to link chemical structure features to biological activity. QSAR involves basically three steps: 1) selection of a training set of bioactive chemicals; 2) selection of appropriate descriptors that can relate more accurately chemical properties to property under investigation; and 3) correlation between structural modifications and changes in property values using statistical techniques.

Logically, 3D-QSAR methods use 3D descriptors in step 2) and the 2D-QSAR methods use topological ones. Because of the large variety of molecular descriptors, the selection of the most adequate ones is crucial for the performance of the QSAR methodology and also depends on the difficulty for modeling the property of interest [17].

3D-QSAR approaches were expected to provide more accurate and reliable models for bioactivity predictions, as they use more complex

3D descriptors. But surprisingly, a wide range of studies have demonstrated that more often than expected 2D-QSAR methods outperform 3D-QSAR [18–22]. This is illustrated in Figure 2, which is also very much in agreement with our internal experience In this respect, there is a need for better understanding the dependency of particular 3D-QSAR methods on conformational changes but also for better selection of 3D descriptors [23]. The higher computational cost of 3D approaches with respect to 2D also makes them less appealing for high-throughput virtual screening of large chemical libraries.

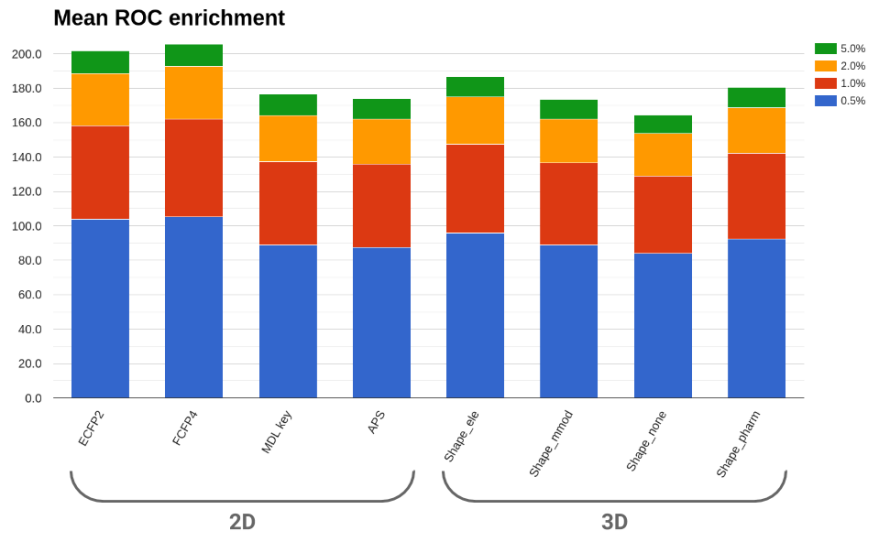

Figure 2. Adapted from [21]
Comparison of the Virtual Screening performances of selected 2D fingerprint and 3D Phase Shape methods using DUD_LIB_VS_1.0 data set [24]. ROC enrichment for false positive fractions (0.5%, 1.0%, 2.0%, 5.0%) is defined as ratio between true positive rate and false positive rate [24].

The definition of the applicability domains in QSAR within chemical space is not less challenging and investigation in this area is very intensive [25–29].

Ligand-based high-throughput virtual screening can be accomplished efficiently using both similarity-based and QSAR methods. Its main objective is to identify a limited set of molecules that are likely to be active on a certain protein target or target profile. These methods return also an estimate of the expected biological affinity of the molecules for the targets of interest on the basis of large reference chemical set for which pharmacological data is available in the public domain (*vide infra*). It became an efficient addition and less expensive alternative to *in vitro* screening in pharmaceutical companies [30–33].

Given a set of molecules of interest, the objective of ligand-based virtual profiling (also known as *in silico* target profiling) is to predict the activities of small molecules across all potential targets covering the available biological space. The combination of *in vitro* and *in silico* target profilings are starting to play an important role in the drug discovery workflow to predict drug efficacy but also anticipate potential on- and off-target liabilities [34] .

All QSAR models are based on known data on compound-target affinities. There are various initiatives that collect, curate and publish large amounts of affinity data, such as ChEMBL [35], DrugBank [36], BindingDB [37], IUPHAR/BPS Guide to PHARMACOLOGY [38]. Pharmaceutical companies also have huge repositories of proprietary pharmacological data acquired from their internal *in vitro* HTS campaigns.

Virtual screening and virtual profiling methods are able to provide a more detailed picture of the interactions between small molecules and their biological targets. This picture can be improved by

incorporating target-related information (tissue specificity, involvement in biological pathways, relationship with adverse reactions and diseases) and molecular structure derived information (structural alerts, tissue distribution, metabolization and excretion models). A wider picture at a systems level can help to better prioritize some candidate molecules and filter those that are likely to produce more side effects or those that are less active or less effective (that affect only one target, the function of which can be compensated by another one).

I.2.2 Polypharmacology

The reductionist approach reflected in the 'one-drug – one-target' paradigm became obsolete at the end of the XXth century. Since then, polypharmacology has been gaining relevance and it is now widely accepted that drugs act on multiple protein targets [39–42]. From the perspective of proteins, their promiscuity had been acknowledged before, specially in enzymes [43–45].

These drug – target promiscuity can be understood as an adequate mechanism for accurate control of the complex biological system, considering the limited number of endogenous ligands and proteins that can interact. Therefore, being able to capture and consider these promiscuities is now absolutely imperative in drug discovery [46].

Information on the off-target interactions of a molecule beyond its primary target can be exploited from two perspectives: 1) increase drug effectiveness by synergistically modulating proteins that play homologous functions; 2) decrease toxicity by designing/optimizing molecules with affinity for less off-targets related to adverse effects. This is a complex task that should take advantage from modern

computational tools to predict drug-target interactions and statistical or prior mechanistic knowledge of the involvement of targets in adverse effects. Thus, in order to develop more effective and safer drugs, both perspectives should be taken into consideration.

Moreover, new integrative computational approaches that combine pharmacological data with safety data for drugs can help gaining a deeper mechanistic understanding of the causal role of certain proteins in serious adverse drug reactions.

Polypharmacology can also assist in recent initiatives for drug repurposing, aiming at identifying a new therapeutical uses for failed drugs in clinical phases, withdrawn drugs due to safety reasons in a particular therapeutic area, as well as good old off-patent drugs. In this respect, the complex nature of cancer makes it an attractive therapeutic area for multi-target drugs. Under this premise, several kinase inhibitors have been developed to strike cancer at more than one relevant protein, e.g. dual inhibitors of PI3K/mTOR [47]. Besides being more effective, they offer also the advantage to overcome possible resistance mechanisms in the corresponding pathways.

The high levels of cross-pharmacology among CNS targets (mainly GPCRs, ion channels, and transporters) make this organ system a difficult target to develop highly selective medicines. On the other hand, this fact may indicate that the natural processes related with such targets share this behaviour and should be treated accordingly. Because of this, single target agents (STAs) have failed the efficacy tests for psychiatric therapies. In contrast, multi-target agents such as Quetiapine have succeeded in providing effective treatment of a

various range of these disorders such as schizophrenia, bipolar disorder and major depressive disorder [48–50]. The role of unique target profiles was considered in the development of drugs like Agomelatine, Vilazodone, Asenapine (Figure 2 [51]) and many others.

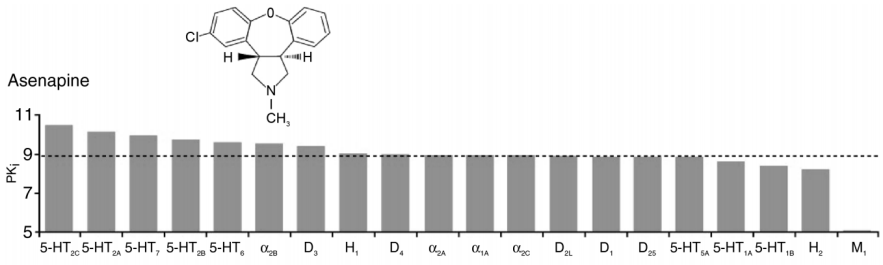

Figure 3. From [51].
Receptor signature of asenapine. The dotted line represents the pK_i for the D_{2L} receptor.

I.2.3 Systems biology approaches for drug design

Systems approaches existed since the origins of drug discovery. A traditional method of finding the appropriate potion for a disease condition and observing the outcome of the mixture consumption directly on human was the first step in a study of complex biological systems. With advances in biochemistry and molecular biology, the methods became more precise, but also more reductionist, obviating the interconnections between the parts of the system isolated by the studies. In order to improve the modern drug design approach it was necessary to change the mentality from the traditional one-drug–one-target principle to incorporate the view that the response mechanisms triggered by the drug action are network processes [46,52,53].

Therefore, the objective of novel systemic approaches is to tackle pathologies and drug actions from the level of biological pathways in cell, intercellular mechanisms that form tissues, organs and the whole organism [54]. The primary requirement to accomplish this objective is the integration of all volume of the data generated from both reductionist and -omics approaches. As Ideker et al. [55] concluded, the integration – in all its senses – plays very important role in systems biology. Foremost, the integration of biological information generated from DNA sequencing, μRNA arrays, genotyping, proteomics, protein interactions, molecular activities and other throughput techniques. The biological knowledge has to be accompanied with computational analytical tools. The interdisciplinary teams of biologists, chemists, technologists, and computational scientists have to be established. And last, but not least, the collaboration between industry and academia would push forward research capabilities.

In the context of drug discovery, one of the implications of the systems biology approach is consideration of ligand multi-targeting properties (discussed in section I.2.2 Polypharmacology) and understanding of the effects that cause drugs on the protein interaction networks [53]. Thus, if only QSAR approaches are applied, it will not provide a full picture for an effective treatment. The information on cellular networks where these targets are involved could improve the strategy by attacking them in the optimal points (Figure 4).

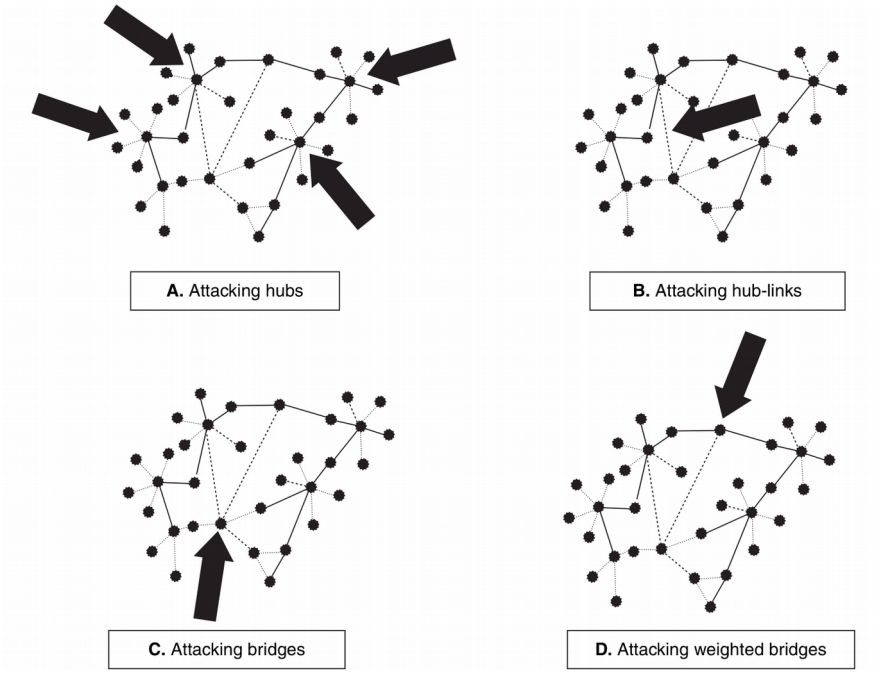

Figure 4. From [56]
Attack scenarios on networks.
A. Attacks on nodes with highest degree (hubs) B. Attacks on 'hub-links' with the highest degree of their endpoints C. Attacks on bridging elements with links having a high betweenness centrality D. Attacks on bridges having links with the highest weighted centrality. Solid and broken lines denote strong and weak links, respectively.

The network configuration data is represented in a number of databases that are collecting, classifying and analysing biological pathways, such as KEGG [57], Pathway Commons [58], WikiPathways [59] or Reactome [60]. This knowledge on interconnections between molecules and proteins can be exploited from different complexity levels – from annotating targets to the pathways in which they are involved, up to simulating all the perturbations that an input molecule could cause.

There are diverse techniques that contribute to the biological systems understanding:

- **Genomics** provides such information as populational and individual variations of DNA [61–63], epigenetic states of DNA [64,65] and differential analysis of mRNA expression patterns in disease states and after pharmacological treatment.

- **Proteomics** measures expression levels and states of proteins in a system of interest. It can give more more direct insights on pathway networks [66] and assist in identification of novel drug targets [67]. Furthermore, chemical proteomics is a promising source of data for drug-target affinities [68,69].

- **Metabolomics** studies variations of endogenous compounds and their metabolites. Metabolic patterns in human plasma in disease state or after a drug treatment is an additional instrument for identification of pathological biomarkers and for assessment of drug effects on a system [70–72] .

Systems pharmacology methods have demonstrated good results in assessing arrhythmias [73] and myocardial infarction [74]. Systems biology approaches were used in a large collaborative study that identified biomarkers for drug-induced liver injury for those cases when *in vivo* models could not provide any direct evidence [75].

I.3 Target-centered systems biology approach

For development of this Thesis, a protein was considered as the central entity of the biological system. It is regarded as a bridge to everything else in the network of multiple interactions.

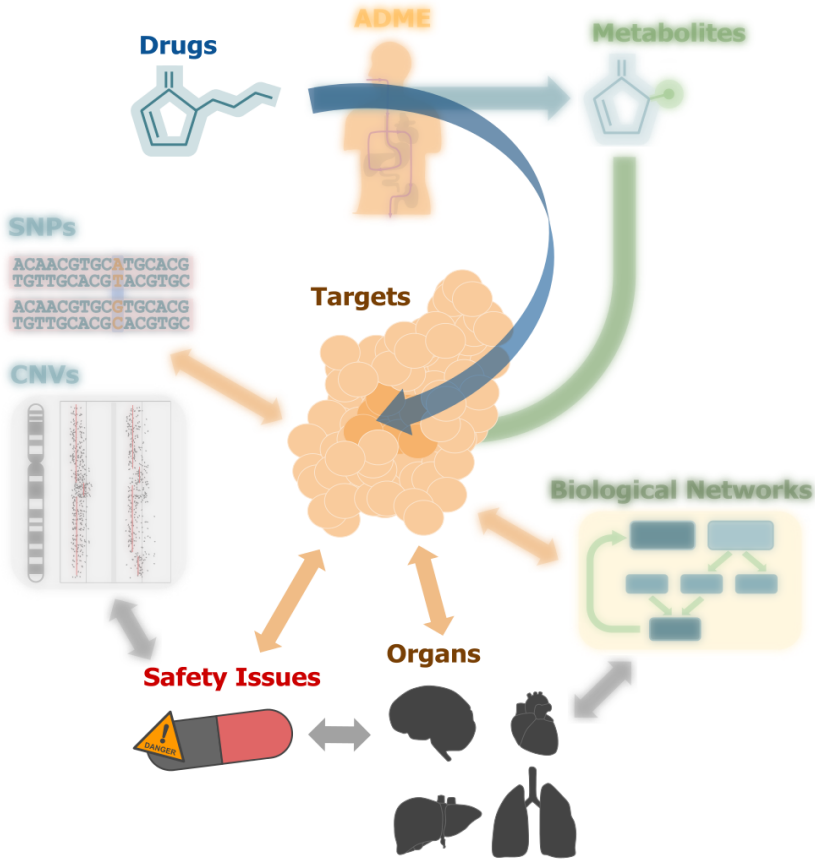

Figure 5
Biological system framework for current PhD research.

This vision of complex relationships between systemic units can be explained schematically (Figure 5): at the system's input is placed the "Drug" entity that becomes a part of the system. The molecules are distributed by all the human body and, eventually, reach different "Targets". It appears here, the first relationship Molecules – Targets

that plays a crucial role in the whole network and must be well characterized. Drug targets, like other human proteins, are expressed at different levels in various anatomical entities, e.g. tissues, organs or organ systems. Therefore, 2 new relationships arise: Target–Organ and Drug–Organ. What is generically called as “ADME” (absorption, distribution, metabolism and excretion) expresses the actions that the system performs on the drug (pharmacokinetics). Among these effects, features the “Metabolite” entity resulting from structural transformation of drugs through the action of chemical reactions that take place in active sites of certain enzymes. The production of these metabolites largely determines the efficacy (and often toxicity) of the medicine. Finally, through the interaction with multiple targets the drug may eventually affect certain “Biological Pathways”, being relevant to various disease states, causing multiple “Side Effects”, and all this in a very miscellaneous manner, according to the presence of single nucleotide polymorphisms (SNPs), copy number variations (CNVs) and epigenetics of each individual in particular. All this significantly increases the number and the type of systemic relationships between the different units, which must be characterized appropriately.

Current thesis will center the attention on the first part of the scheme (non-blurred entities on the Figure 5): the exploitation of drug-target interactions for different strategies in drug design; the anatomical profiling of drug-targets and, consecutively, anatomical drug projection.

I.3.1 Target deconvolution

Drug targets identification and verification are essential tasks in rational drug design process. The discovery of proteins, that can be inhibited or activated by small molecules in certain disease states where their functions are altered, was accelerated by novel genomics and proteomics techniques [76,77]. But still, many traditional drugs have not any targets assigned, lacking the mechanistic explanation of their action [78].

On the other hand, as already discussed, reductionist target-based drug discovery, that was based on screening large chemical libraries against a potential target, did not give expected increase in productivity. Therefore, pharmaceutical industry and academia are returning to, abandoned in 1990s, phenotypical assays that can provide direct information on drug effects in biological systems of different levels: cell cultures, tissues, isolated organs or model organisms [79].

One of the most valuable features of phenotype-based approaches is their ability to detect active molecules in close to real environment of complex system. Phenotypic screenings can assist in identification of new proteins and pathways that play role in a given phenotypic output, understanding underlying mechanisms and further optimization of active compounds. There are several experimental techniques to determine targets that interact with active molecules discovered in phenotypic screens [80]. For example, Affinity chromatography aided to deconvolute targets for drugs like Roscovitine, Melanogenin, Diazonamide A, Indomethacin and others. There are many alternatives e.g. three-hybrid systems, Phage

display, mRNA display, Protein microarrays, Reverse transfected cell microarrays, Biochemical suppression. Furthermore, *in silico* approaches are able to accelerate the process by providing target predictions based on available models. Because these computational analyses generate predictions, direct proof of compound-protein interaction and its functional consequences is still necessary.

Frequently, the molecule's active outcome from phenotypic screening can not be explained only by one target, and polypharmacology comes to the forefront again. In case of kinase inhibitors, for example, an unexpected interactions with off-target kinases were recently discovered, even for well-known, previously thought to be specific, kinase inhibitors [81]. This discovery casted doubt on the assumed mode of action of numerous kinase inhibitors, as it was difficult to draw any conclusion on which pathways were interfered by these molecules. Using mathematical approach named elastic net regularization, that regresses a single variable (quantified cell migration) against a set of predictor variables (the activities of individual kinases), *a posteriori,* it was possible to identify informative kinases that are responsible for the phenotype [82].

The complete activity matrix of more than 500 kinases permit the researchers to deconvolute those targets that are related to the outcome of their interest, for example primary chronic lymphocytic leukemia [83].

In context of drug development, target deconvolution is important in follow-up studies, giving assistance to medicinal chemists. Furthermore, identifying both the primary target and other off-

targets, that could produce side effects, allows optimization of small-molecule effectiveness.

I.3.2 Chemical library design

A standard lead-searching process implies application of combinatorial chemistry and high-throughput screening. The latter requires availability of a large compound collection assembled from synthetic molecules, combinatorial libraries and natural products. A success of a high throughput screening relies on diversity of the chemical library. Thus, the selection of molecules that do not contain redundant properties is important in order to reduce time and financial wastage [84,85].

A classical chemistry-based approach for diversity selection for a biological screening is based on the previously mentioned Similar Property Principle formulated by Johnson and Maggiora [7] that implies, in this case, that in order to avoid inclusion of biologically-similar molecules into chemical library, they must be compared and separated using chemical descriptors. Hans Matter evaluated different 1D, 2D and 3D descriptors and concluded that any of them is more effective than a random selection and 2D fingerprint-based descriptors showed the best results [86].

Another approach proposed by Dixon and Villar [87] starts from the premise that there are compounds which appear to be very similar structurally but manifest different activities to a given target. On the other hand, some targets also exhibit promiscuity, interacting with molecules that are different structurally. The authors suggest the utilization of affinity fingerprints (the set of activities measured for a single molecule across the biological space), thus overcoming the

difficulties associated with selecting compounds on the basis of structure alone. This approximation may be considered as "biological descriptors" and may become a powerful tool for rational selection. Moreover, the biological information encoded by affinity fingerprints is relevant to any target that share similar binding site to any protein from the screening panel.

Similar approaches, based on biological activity profiles, are used to compare compounds, for example, high-throughput screening fingerprints" (HTS-FPs) described in [88].

The methods mentioned above are based on experimental gathered bioactivity profiles of the molecules, but there is an opportunity for *in silico* pharmacological profiles to expand the biological coverage, using similarity-based predictions.

I.3.3 Predictive Drug Safety

The anticipation of drug toxicity is an important task both for public health and for pharmaceutical companies. A number of studies and meta-analyses have demonstrated that drug-related hospitalizations (DRH) in different countries account for between 2.4% and 6.2% of all medical admissions. In case of Germany, the annual direct costs were estimated up to 400 million euros. [89]. As for impact in pharmaceutical industry, the project closures in clinical trials account for up to 44%, in case of AstraZeneca [3], despite the overall number for industry is slightly smaller.

The drug safety assessment started to evolve since the nineteenth century, when new drugs were studied initially on man with little or no previous study in laboratory animals. By the end of the nineteenth

century, predictive toxicology had progressed to the point that Dreser, announcing the discovery of acetylsalicylic acid, reported its toxicity on fish and frog. Since then, and until three decades ago, most of the predictive toxicology studies were based on preclinical trials using laboratory animals [90].

Since the second half of the 20th century, pharmacologists started to be aware of problems related to the extrapolation of the *in vivo* toxicity studies results from animals to humans [91,92]. The scientific community realized the need of improvement in the capacity to extrapolate from animal models to humans and its limitations [93,94].

The dissimilarities in anatomy, physiology and metabolism were discussed by various authors [93,95–97]. These studies provided valuable starting points for anticipating the probability of similar response between laboratory animals and humans caused by a compound of interest.

In 2000 International Life Sciences Institute (ILSI) coordinated the compilation and analysis of data on toxicity of drugs in development, provided by twelve pharmaceutical companies. The result of this first collaboration was reported in [98]. The collaborators found that the nonrodent species (dog and primate) have a higher ratio of positive concordance *vs.* non-concordance than rodents did (Figure 6) and that the frequency of positive concordance varied depending on organ affected (Figure 7). Hematological, gastrointestinal, and cardiovascular toxicities showed the highest values and the lowest scores were for cutaneous and liver.

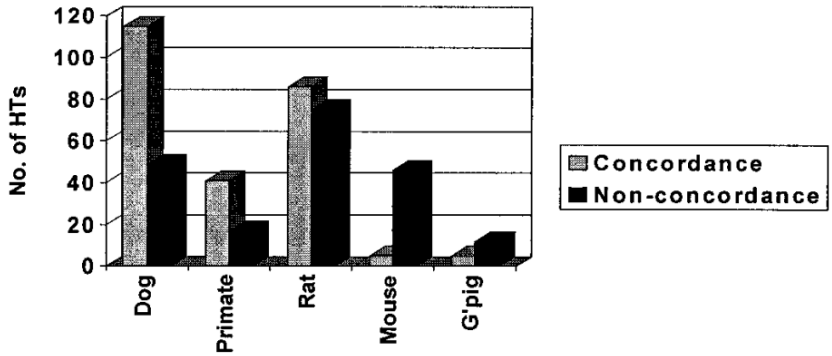

Figure 6. From [98]
Concordance rates *versus* species. (HT = Human Toxicities)

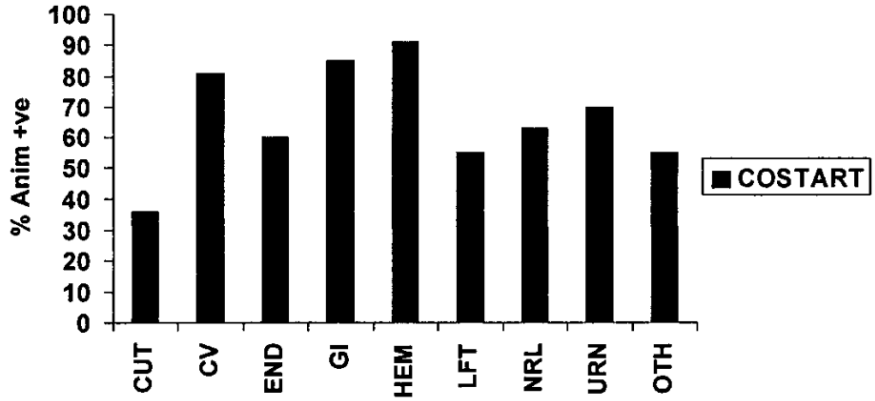

Figure 7. From [98]
Animal concordance by COSTART category. (+ve = positive).
CUT - cutaneous, CV - cardiovascular, END - endocrine, GI - gastrointestina, HEM - hematological, LFT - liver function test, NRL - neurological, URN-urinary, OTH – other

Because of the low value of animal testing as predictive modality for human response to drugs [99], high costs of such experiments and arising ethical issues [100], the demand of advanced techniques emerges in the end of the 20th century. The development of *in vitro* phenotypic experiments (previous and complementary to *in vivo*

studies) to evaluate the toxicity of chemicals and drugs has thus become increasingly important.

The most direct assessment for harmful effects of a molecule on a cell is approached by *in vitro* cytotoxicity tests. Cytotoxicity is defined as the potential of a compound to induce cell death. Information on cytotoxicity can be used as an indicator of acute toxic effects *in vivo* [101].

From the perspective of regulator organs, *in vitro* assessments were only required for genetic toxicology, until recently. Since 2005, cardiac safety pharmacology and phototoxicity tests also became mandatory [102].

Some examples of *in vitro* model systems were described in [102,103], and can be summarized as followed:

- **Cutaneous** and **ocular** toxicity models. Demonstrate good correlation with *in vivo* outcome.
- **Hepatotoxicity** models. Have frequently been described as species-specific and an idiosyncratic phenomenon, much work has been dedicated to establishing predictive models using human hepatocyte cultures and cell lines derived from the human liver.
- **Genotoxicity** models. The most used Ames test has high sensitivity, but low specificity. Carcinogenicity models, in general, have room to improve.
- **Cardiotoxicity:** Screenit system [104] was found to be highly predictive for the proarrhythmic potential.

With an increasing -omics and systems biology knowledge the number of proteins associated with adverse drug reactions is growing constantly. It was estimated that around 75% of all adverse drug reactions (ADRs) are dose-dependent and can be anticipated on the basis of the pharmacology profiles of the candidate molecule [105]. Off-target activity is frequent cause of side-effects in animal models and clinical studies. Therefore, detailed characterization of secondary pharmacology profiles of drug candidates at the beginning of the drug discovery workflow could help to reduce the incidence of type A adverse reactions (ones that are not immunologic nor idiosyncratic).

The first obligatory *in vitro* affinity assay that became required by regulatory authorities (both FDA and MEA) is one that measures the effects of candidate drugs on the ionic current of hERG (human voltage-gated potassium channel subfamily H member 2, KCNH2). At the moment, current regulatory guidances do not give recommendations on which proteins should be included in an in vitro pharmacological profiling panel. Nevertheless, most pharmaceutical companies already perform this testing early in drug discovery to reduce attrition and to facilitate better prediction of side effects in the later stages of drug discovery and development.

In 2012, for the first time, four major pharmaceutical companies (AstraZeneca, GlaxoSmithKline, Novartis and Pfizer) published the collaborative study, in which they discuss rationale, strategies and methodologies for *in vitro* pharmacological profiling [105]. 44 targets (25 were used by three out of four companies and 19 by all of them) were proposed to be considered as a minimal panel that should

provide a wide early anticipation of the toxicity effect of a compound or chemical series.

Another advantage of *in vitro* pharmaceutical profiling is that it can be exploited both for off-target related toxicity alerts and for hit optimization against primary targets on the same screening experiment in parallel.

The data generated during these 5–10 years in compound screening assays against the protein sets corresponding to each company, is a gold mine for computational and systems biology analysis, which may provide better understanding of mechanisms of related ADRs, combining bioactivity and phenotypic information obtained in this process.

Summarizing, in the course of the evolution in science and in research methodologies, the more high throughput and cost-effective approaches were developed. Every innovation was applied to the existing clinical and preclinical toxicology tests. All started with direct human trials; later on laboratory animals were added to the previous step in order to decrease human costs; consequently *in vitro* phenotypic assays were introduces to the stage previous to the animal testing in order to reduce costs, timescale and improve predictability; afterwards *in vitro* pharmacological profiling was added to precede other tests with aim to test much broader chemical series in less time and prioritize optimal candidates to pass to the next step. Finally, *in silico* predictions of target and safety profiles may be used as the very first step in order to optimize costs and anticipate insights of the mechanism of action on the human body complex system.

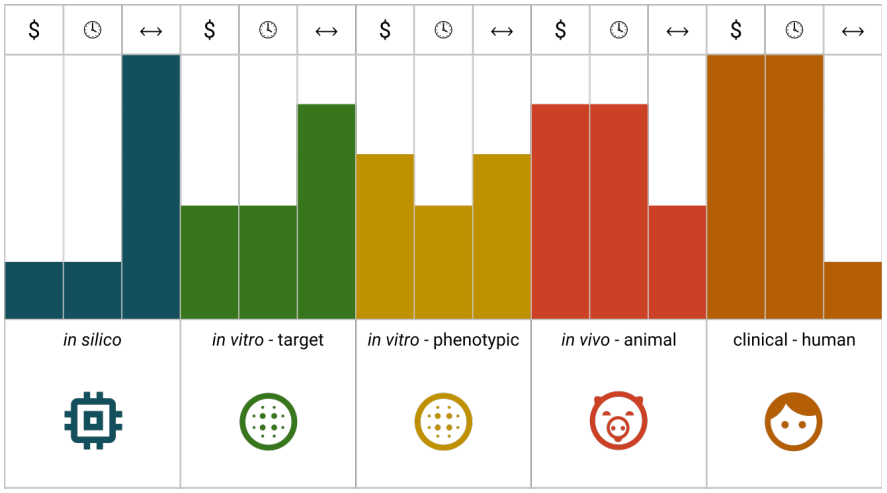

Figure 8
Overview of different toxicity prediction approaches and their comparison in terms of time, cost, and throughput

I.4 Particularities of PhD research in a computational biotech company

I was offered the opportunity to do the Thesis I am now presenting at the Chemotargets company thanks to a "TALENT-empresa" fellowship from the *Generalitat de Catalunya*. The objective of this type of grants was to promote R&D in businesses located in Catalonia and foster the development of business aspects into the training of new PhD researchers.

Chemotargets S.L. is a computationally-oriented biotech company focused on the drug discovery and development area. Located in Barcelona, it was founded on March 2006 as a spin-off initiative from the Chemogenomics Laboratory (now called Systems Pharmacology group) under the auspices of the Municipal Institute of Medical Research (IMIM-Hospital del Mar). The founder and president of the company, Dr. Jordi Mestres, is also the head of the research group on Systems Pharmacology of the Research Program on Biomedical Informatics (GRIB) of the IMIM/UPF.

R&D in a company is expected to bring profit in the medium to long term. Chemotargets was interested in incorporating a PhD student with a good background on molecular biology and bioinformatics of the drug, as it would allow the company, on the one hand, increase the effort in the field of research on the relationships between the different biological system units and, on the other hand, transfer this methodology to the market, either by licensing to third parties or integrating it into a new software that was at the time in development at Chemotargets.

Accordingly, the present Thesis will be focused on the practical implementation of the Chemotargets methodologies from the point of view of systems biology and also on the development of new methodologies of current interest for pharma industry, our main client base.

Overall, it's been a great experience that has allowed me to do research with a business mind, so research lines are not only valued by their positive impact in science but also on their potential market value based on specific requirements from pharmaceutical customers. Publishing a scientific article makes you feel proud of your work and gives you a nice feeling, but knowing that the software you have developed is currently being used by 2 of the top-10 pharmaceutical companies to predict the pharmacological and safety profiles of their small molecules, is something quite special. And I am happy that my Thesis reflects both aspects.

I.5 References

1. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov. 2010;9: 203–214.
2. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. Nat Biotechnol. 2014;32: 40–51.
3. Cook D, Brown D, Alexander R, March R, Morgan P, Satterthwaite G, et al. Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. Nat Rev Drug Discov. 2014;13: 419–431.
4. Bennani YL. Drug discovery in the next decade: innovation needed ASAP. Drug Discov Today. 2011;16: 779–792.
5. Sams-Dodd F. Target-based drug discovery: is something wrong? Drug Discov Today. 2005;10: 139–147.
6. Sliwoski G, Kothiwale S, Meiler J, Lowe EW Jr. Computational Methods in Drug Discovery. Pharmacol Rev. American Society for Pharmacology and Experimental Therapeutics; 2014;66: 334–395.
7. Johnson MA, Maggiora GM. Concepts and applications of molecular similarity. Wiley; 1990.
8. Wagner G, Hyberts SG, Havel TF. NMR structure determination in solution: a critique and comparison with X-ray crystallography. Annu Rev Biophys Biomol Struct. 1992;21: 167–198.
9. Baker D, Sali A. Protein structure prediction and structural genomics. Science. 2001;294: 93–96.

10. Floudas CA, Fung HK, McAllister SR, Mönnigmann M, Rajgaria R. Advances in protein structure prediction and de novo protein design: A review. Chem Eng Sci. 2006;61: 966–988.

11. Kelley LA, Sternberg MJE. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc. 2009;4: 363–371.

12. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc. 2010;5: 725–738.

13. Ripphausen P, Nisius B, Peltason L, Bajorath J. Quo vadis, virtual screening? A comprehensive survey of prospective applications. J Med Chem. 2010;53: 8461–8467.

14. Stumpfe D, Dagmar S, Peter R, Jürgen B. Virtual compound screening in drug discovery. Future Med Chem. 2012;4: 593–602.

15. Xue L, Bajorath J. Molecular descriptors in chemoinformatics, computational combinatorial chemistry, and virtual screening. Comb Chem High Throughput Screen. 2000;3: 363–372.

16. Todeschini R, Consonni V. Handbook of molecular descriptors. John Wiley & Sons; 2008.

17. Ekins S, Mestres J, Testa B. In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. Br J Pharmacol. 2007;152: 9–20.

18. Perkins R, Fang H, Tong W, Welsh WJ. Quantitative structure-activity relationship methods: Perspectives on drug discovery and toxicology. Environ Toxicol Chem. Wiley Periodicals, Inc.; 2003;22: 1666–1679.

19. Chohan KK, Paine SW, Waters NJ. Quantitative structure activity relationships in drug metabolism. Curr Top Med Chem. 2006;6: 1569–1578.

20. Venkatraman V, Pérez-Nueno VI, Mavridis L, Ritchie DW. Comprehensive comparison of ligand-based virtual screening tools against the DUD data set reveals limitations of current 3D methods. J Chem Inf Model. 2010;50: 2079–2093.

21. Hu G, Kuang G, Xiao W, Li W, Liu G, Tang Y. Performance evaluation of 2D fingerprint and 3D shape similarity methods in virtual screening. J Chem Inf Model. 2012;52: 1103–1113.

22. Goodarzi M, Heyden YV, Funar-Timofei S. Towards better understanding of feature-selection or reduction techniques for Quantitative Structure–Activity Relationship models. Trends Analyt Chem. 2013;42: 49–63.

23. Hechinger M, Leonhard K, Marquardt W. What is wrong with quantitative structure-property relations models based on three-dimensional descriptors? J Chem Inf Model. 2012;52: 1984–1993.

24. Jahn A, Hinselmann G, Fechner N, Zell A. Optimal assignment methods for ligand-based virtual screening. J Cheminform. 2009;1: 14.

25. Dimitrov S, Dimitrova G, Pavlov T, Dimitrova N, Patlewicz G, Niemela J, et al. A stepwise approach for defining the applicability domain of SAR and QSAR models. J Chem Inf Model. 2005;45: 839–849.

26. Ekins S, Mestres J, Testa B. In silico pharmacology for drug discovery: applications to targets and beyond. Br J Pharmacol. 2007;152: 21–37.

27. Weaver S, Gleeson MP. The importance of the domain of applicability in QSAR modeling. J Mol Graph Model. 2008;26: 1315–1326.

28. Dragos H, Gilles M, Alexandre V. Predicting the Predictability: A Unified Approach to the Applicability Domain Problem of QSAR Models. J Chem Inf Model. 2009;49: 1762–1776.

29. Sahigara F, Mansouri K, Ballabio D, Mauri A, Consonni V, Todeschini R. Comparison of Different Approaches to Define the Applicability Domain of QSAR Models. Molecules. 2012;17: 4791–4810.

30. Stahura FL, Bajorath J. Virtual screening methods that complement HTS. Comb Chem High Throughput Screen. 2004;7: 259–269.

31. Willett P. Similarity-based virtual screening using 2D fingerprints. Drug Discov Today. 2006;11: 1046–1053.

32. Gregori-Puigjane E, Mestres J. A Ligand-Based Approach to Mining the Chemogenomic Space of Drugs. CCHTS. 2008;11: 669–676.

33. Geppert H, Vogt M, Bajorath J. Current trends in ligand-based virtual screening: molecular representations, data mining methods, new application areas, and performance evaluation. J Chem Inf Model. 2010;50: 205–216.

34. Schmidt F, Matter H, Hessler G, Czich A. Predictive in silico off-target profiling in drug discovery. Future Med Chem. 2014;6: 295–317.

35. Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M, et al. The ChEMBL bioactivity database: an update. Nucleic Acids Res. 2014;42: D1083–90.

36. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 2006;34: D668–72.

37. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. Nucleic Acids Res. 2007;35: D198–201.

38. Southan C, Sharman JL, Benson HE, Faccenda E, Pawson AJ, Alexander SPH, et al. The IUPHAR/BPS Guide to PHARMACOLOGY in 2016: towards curated quantitative interactions between 1300 protein targets and 6000 ligands. Nucleic Acids Res. 2015; doi:10.1093/nar/gkv1037

39. Morphy R, Kay C, Rankovic Z. From magic bullets to designed multiple ligands. Drug Discov Today. 2004;9: 641–651.

40. Jalencas X, Mestres J. On the origins of drug polypharmacology. Med Chem Commun. The Royal Society of Chemistry; 2012;4: 80–87.

41. Lu J-J, Pan W, Hu Y-J, Wang Y-T. Multi-target drugs: the trend of drug research and development. PLoS One. 2012;7: e40262.

42. Anighoro A, Andrew A, Jürgen B, Giulio R. Polypharmacology: Challenges and Opportunities in Drug Discovery. J Med Chem. 2014;57: 7874–7887.

43. Schreiber G, Keating AE. Protein binding specificity versus promiscuity. Curr Opin Struct Biol. 2011;21: 50–61.

44. Nobeli I, Favia AD, Thornton JM. Protein promiscuity and its implications for biotechnology. Nat Biotechnol. 2009;27: 157–167.

45. O'Brien PJ, Herschlag D. Catalytic promiscuity and the evolution of new enzymatic activities. Chem Biol. 1999;6: R91–R105.

46. Brown JB, Okuno Y. Systems biology and systems chemistry: new directions for drug discovery. Chem Biol. 2012;19: 23–28.

47. Karlsson A, García-Echeverría C. Chapter 13. Identification and Optimization of Dual PI3K/mTOR Inhibitors. In: Morphy JR, Harris CJ, editors. Designing Multi-Target Drugs. Royal Society of Chemistry; 2012. pp. 206–220.

48. Shahid M, Mohammed S. Chapter 2. Clinical Need and Rationale for Multi-Target Drugs in Psychiatry. RSC Drug Discovery. 2012. pp. 14–31.

49. Roth BL, Sheffler DJ, Kroeze WK. Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. Nat Rev Drug Discov. 2004;3: 353–359.

50. Roth BL, Sheffler D, Potkin SG. Atypical antipsychotic drug actions: unitary or multiple mechanisms for “atypicality”? Clin Neurosci Res. 2003;3: 108–117.

51. Shahid M, Walker GB, Zorn SH, Wong EHF. Asenapine: a novel psychopharmacologic agent with a unique human receptor signature. J Psychopharmacol. 2009;23: 65–73.

52. Van der Greef J, McBurney RN. Rescuing drug discovery: in vivo systems pathology and systems pharmacology. Nat Rev Drug Discov. Nature Publishing Group; 2005;4: 961–967.

53. Zhao S, Iyengar R. Systems pharmacology: network analysis to identify multiscale mechanisms of drug action. Annu Rev Pharmacol Toxicol. 2012;52: 505–521.

54. Butcher EC, Berg EL, Kunkel EJ. Systems biology in drug discovery. Nat Biotechnol. 2004;22: 1253–1259.

55. Ideker T, Galitski T, Hood L. A new approach to decoding life: systems biology. Annu Rev Genomics Hum Genet. 2001;2: 343–372.

56. Korcsmáros T, Szalay MS, Böde C, Kovács IA, Csermely P. How to design multi-target drugs. Expert Opin Drug Discov. 2007;2: 799–808.

57. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28: 27–30.

58. Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O, Anwar N, et al. Pathway Commons, a web resource for biological pathway data. Nucleic Acids Res. 2011;39: D685–90.

59. Kelder T, van Iersel MP, Hanspers K, Kutmon M, Conklin BR, Evelo CT, et al. WikiPathways: building research communities on biological pathways. Nucleic Acids Res. 2012;40: D1301–7.

60. Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. The Reactome pathway knowledgebase. Nucleic Acids Res. 2014;42: D472–7.

61. Cooper RA, Ferdig MT, Su X-Z, Ursos LMB, Mu J, Nomura T, et al. Alternative Mutations at Position 76 of the Vacuolar Transmembrane Protein PfCRT Are Associated with Chloroquine Resistance and Unique Stereospecific Quinine and Quinidine Responses in Plasmodium falciparum. Mol Pharmacol. ASPET; 2002;61: 35–42.

62. Rost S, Fregin A, Ivaskevicius V, Conzelmann E, Hörtnagel K, Pelz H-J, et al. Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2. Nature. 2004;427: 537–541.

63. Thompson JF, Man M, Johnson KJ, Wood LS, Lira ME, Lloyd DB, et al. An association study of 43 SNPs in 16 candidate genes with atorvastatin response. Pharmacogenomics J. 2005;5: 352–358.

64. Esteller M. Epigenetics in cancer. N Engl J Med. 2008;358: 1148–1159.

65. Narayan P, Dragunow M. Pharmacology of epigenetics in brain disorders. Br J Pharmacol. 2010;159: 285–303.

66. Langley SR, Dwyer J, Drozdov I, Yin X, Mayr M. Proteomics: from single molecules to biological pathways. Cardiovasc Res. 2013;97: 612–622.

67. Gottfries J, Sjögren M, Holmberg B, Rosengren L, Davidsson P, Blennow K. Proteomics for drug target discovery. Chemometrics Intellig Lab Syst. 2004;73: 47–53.

68. Katayama H, Oda Y. Chemical proteomics for drug discovery based on compound-immobilized affinity chromatography. J Chromatogr B Analyt Technol Biomed Life Sci. 2007;855: 21–27.

69. Rix U, Superti-Furga G. Target profiling of small molecules by chemical proteomics. Nat Chem Biol. 2009;5: 616–624.

70. Kaddurah-Daouk R, Kristal BS, Weinshilboum RM. Metabolomics: a global biochemical approach to drug response and disease. Annu Rev Pharmacol Toxicol. 2008;48: 653–683.

71. Powers R. NMR metabolomics and drug discovery. Magn Reson Chem. 2009;47 Suppl 1: S2–11.

72. Beger RD, Sun J, Schnackenberg LK. Metabolomics approaches for discovering biomarkers of drug-induced hepatotoxicity and nephrotoxicity. Toxicol Appl Pharmacol. 2010;243: 154–166.

73. Berger SI, Ma'ayan A, Iyengar R. Systems pharmacology of arrhythmias. Sci Signal. 2010;3: ra30.

74. Azuaje FJ, Zhang L, Devaux Y, Wagner DR. Drug-target network in myocardial infarction reveals multiple side effects of unrelated drugs. Sci Rep. 2011;1: 52.

75. McBurney RN, Hines WM, Tungeln LS Von, Schnackenberg LK, Beger RD, Moland CL, et al. The liver toxicity biomarker study: phase I design and preliminary results. Toxicol Pathol. 2009;37: 52–64.

76. Kramer R, Cohen D. Functional genomics to new drug targets. Nat Rev Drug Discov. 2004;3: 965–972.

77. Bakheet TM, Doig AJ. Properties and identification of human protein drug targets. Bioinformatics. 2009;25: 451–457.

78. Imming P, Sinning C, Meyer A. Drugs, their targets and the nature and number of drug targets. Nat Rev Drug Discov. 2006;5: 821–834.

79. Terstappen GC, Schlüpen C, Raggiaschi R, Gaviraghi G. Target deconvolution strategies in drug discovery. Nat Rev Drug Discov. 2007;6: 891–903.

80. Lee J, Bogyo M. Target deconvolution techniques in modern phenotypic profiling. Curr Opin Chem Biol. 2013;17: 118–126.

81. Anastassiadis T, Deacon SW, Devarajan K, Ma H, Peterson JR. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. Nat Biotechnol. 2011;29: 1039–1045.

82. Gujral TS, Peshkin L, Kirschner MW. Exploiting polypharmacology for drug target deconvolution. Proc Natl Acad Sci U S A. 2014;111: 5048–5053.

83. Kruse U, Pallasch CP, Bantscheff M, Eberhard D, Frenzel L, Ghidelli S, et al. Chemoproteomics-based kinome profiling and target deconvolution of clinical multi-kinase inhibitors in primary chronic lymphocytic leukemia cells. Leukemia. 2010;25: 89–100.

84. Ferguson AM, Patterson DE, Garr CD, Underiner TL. Designing Chemical Libraries for Lead Discovery. J Biomol Screen. 1996;1: 65–73.

85. Ellman JA. Design, Synthesis, and Evaluation of Small-Molecule Libraries. Acc Chem Res. 1996;29: 132–143.

86. Matter H. Selecting optimally diverse compounds from structure databases: a validation study of two-dimensional and three-dimensional molecular descriptors. J Med Chem. 1997;40: 1219–1229.

87. Dixon SL, Villar HO. Bioactive diversity and screening library selection via affinity fingerprinting. J Chem Inf Comput Sci. 1998;38: 1192–1203.

88. Petrone PM, Simms B, Nigsch F, Lounkine E, Kutchukian P, Cornett A, et al. Rethinking Molecular Similarity: Comparing Compounds on the Basis of Biological Activity. ACS Chem Biol. 2012;7: 1399–1409.

89. Schneeweiss S, Hasford J, Göttler M, Hoffmann A, Riethling A-K, Avorn J. Admissions caused by adverse drug events to internal medicine and emergency departments in hospitals: a longitudinal population-based study. Eur J Clin Pharmacol. 2002;58: 285–291.

90. Boyd EM. Predictive drug toxicity. Assessment of drug safety before human use. Can Med Assoc J. 1968;98: 278–293.

91. Litchfield JT Jr. Symposium on clinical drug evaluation and human pharmacology. XVI. Evaluation of the safety of new drugs by means of tests in animals. Clin Pharmacol Ther. 1962;3: 665–672.

92. Rall DP. Difficulties in extrapolating the results of toxicity studies in laboratory animals to man. Environ Res. 1969;2: 360–367.

93. Calabrese EJ. Suitability of Animal Models for Predictive Toxicology: Theoretical and Practical Considerations. Drug Metab Rev. 1984;15: 505–523.

94. Knight A. Systematic reviews of animal experiments demonstrate poor human clinical and toxicological utility. Altern Lab Anim. 2007;35: 641–659.

95. Dixon RL. Problems in extrapolating toxicity data for laboratory animals to man. Environ Health Perspect. 1976;13: 43–50.

96. Heywood R. Target organ toxicity. Toxicol Lett. 1981;8: 349–358.

97. Garattini S. Toxic effects of chemicals: difficulties in extrapolating data from animals to man. Crit Rev Toxicol. 1985;16: 1–29.

98. Olson H, Betton G, Robinson D, Thomas K, Monro A, Kolaja G, et al. Concordance of the toxicity of pharmaceuticals in humans and in animals. Regul Toxicol Pharmacol. 2000;32: 56–67.

99. Greek R, Menache A. Systematic reviews of animal models: methodology versus epistemology. Int J Med Sci. 2013;10: 206–221.

100. Shanks N, Greek R. Experimental use of nonhuman primates is not a simple problem. Nat Med. 2008;14: 1012; discussion 1012–3.

101. Eisenbrand G, Pool-Zobel B, Baker V, Balls M, Blaauboer BJ, Boobis A, et al. Methods of in vitro toxicology. Food Chem Toxicol. 2002;40: 193–236.

102. Suter W. Predictive value of in vitro safety studies. Curr Opin Chem Biol. 2006;10: 362–366.

103. Davila JC, Rodriguez RJ, Melchert RB, Acosta D Jr. Predictive value of in vitro model systems in toxicology. Annu Rev Pharmacol Toxicol. 1998;38: 63–96.

104. Valentin J-P, Hoffmann P, De Clerck F, Hammond TG, Hondeghem L. Review of the predictive value of the Langendorff heart model (Screenit system) in assessing the proarrhythmic potential of drugs. J Pharmacol Toxicol Methods. 2004;49: 171–181.

105. Bowes J, Brown AJ, Hamon J, Jarolimek W, Sridhar A, Waldron G, et al. Reducing safety-related drug attrition: the use of in vitro pharmacological profiling. Nat Rev Drug Discov. 2012;11: 909–922.

Part II: Objectives

The main purpose of the current PhD Thesis was to perform research, development and application of novel integrative systems approaches to drug discovery. Specific objectives are listed below:

1. To establish the framework of research in the context of an interconnected system that can be perturbed by a drug and apply it for large-scale predictive drug safety
2. To extend the applicability domain of systems approaches to deconvolution of disease-relevant targets, design of chemical libraries with wide biological coverage, and identification of protein hazards for main safety endpoints
3. To contribute to the safety module of CT-link by adding the possibility of projecting the anatomical profile of novel bioactive small molecules
4. To design, develop, and manage the commercial distribution of CT-link, Chemotargets' software to predict the pharmacological and safety profile of small molecules

The first objective is described in the Introduction section and, in part, in the **Chapter III.4** (Large-Scale Predictive Drug Safety). The second objective was the most extensive and is detailed in **Chapters III.1–3** wherein different in silico target-centred approaches were applied for identification of cancer-related proteins, compilation of diverse chemical library, and prediction of target-related safety events. The third objective was accomplished by adding the new anatomical "Link" to the framework and is described in **Chapter III.5**. And the fourth objective – the most practical one – is detailed in **Chapter III.6**.

Part III: Results

Flachner B, Lörincz Z, Carotti A, Nicolotti O, Kuchipudi P, Remez N, Sanz F, Tóvári J, Szabó MJ, Bertók B, Cseh S, Mestres J, Dormán G. A chemocentric approach to the identification of cancer targets. PLoS One. 2012;7(4):e35582. doi: 10.1371/journal.pone.0035582.

Horvath D, Lisurek M, Rupp B, Kühne R, Specker E, von Kries J, Rognan D, Andersson CD, Almqvist F, Elofsson M, Enqvist PA, Gustavsson AL, Remez N, Mestres J, Marcou G, Varnek A, Hibert M, Quintana J, Frank R. Design of a general-purpose European compound screening library for EU-OPENSCREEN. ChemMedChem. 2014 Oct;9(10):2309-26. doi: 10.1002/cmdc.201402126.

Schmidt F, Amberg A, Mulliner D, Stolte M, Matter H, Hessler G, Dietrich A, Remez N, Vidal D, Mestres J, Czich A. Computational prediction of off-target related safety liabilities of molecules: Cardiotoxicity, hepatotoxicity and reproductive toxicity. Toxicology Letters, 229, Suppl, 10 September 2014, p. S164, ISSN 0378-4274, http://dx.doi.org/10.1016/j.toxlet.2014.06.564.

Schmidt F, Amberg A, Mulliner D, Stolte M, Brennan R, Matter H, Hessler G, Dietrich A, García R, Remez N, Vidal D, Mestres J, Czich A. Computational prediction of off-target related risks of molecules: Cardiotoxicity, hepatotoxicity and reproductive toxicity [póster] [Consultable a: http://grib.imim.es/media/upload/pdf//1409_bestposterchemotargets_red_editora_5_281_1.pdf] [Data consulta: 28.10.2016]

*ESTIV Best Poster Award at the Eurotox2014

Garcia-Serna R, Vidal D, Remez N, Mestres J. Large-Scale Predictive Drug Safety: From Structural Alerts to Biological Mechanisms. Chem Res Toxicol. 2015 Oct 19;28(10):1875-87. doi: 10.1021/acs.chemrestox.5b00260

Remez N, Garcia-Serna R, Vidal D, Mestres J. The In Vitro Pharmacological Profile of Drugs as a Proxy Indicator of Potential In Vivo Organ Toxicities. Chem Res Toxicol. 2016 Apr 18;29(4):637-48. doi: 10.1021/acs.chemrestox.5b00470

III.6 CT-link distribution package report

Nikita Remez

Introduction

Recent changes on the R&D pharmaceutical market increased a demand for an integrative computational drug discovery platform. Chemotargets had integrated various in-house methods in the manner of “Links” between biological systems entities, starting from “TLINK” (Target), advancing to “SLINK” (Safety), mentioned in current work “ALINK” (Anatomic) and more, all together known as CT-link.

Initially, all the programs at Chemotargets are developed for the internal use and are applied for external or collaborative projects. The methodologies are validated continuously by experimental data and retrospective concordance tests.

Current software design approach gives a possibility to dynamically adapt and reuse the code, but requires a controlled environment for development and implementation. There are several library dependencies that are needed for the correct function of Chemotargets software.

Therefore, the programming language of preference at Chemotargets is Python [1,2] (with some parts of resource-demanding code written in C [3]). Moreover, python has various compatible libraries for molecular processing like OpenBabel and RDKit [4,5]. As an interpreted language it has its limitations e.g. difficulty to distribute the closed-source programs.

Definitely, there was a need to prepare an easily deployable package for a broad range of hardware, that could be scaled from a personal workstation to a corporative server. At the same time, the code had to be protected and the execution control implemented.

Distribution model

The first attempt of distribution by third-party company did not give expected results and had some difficulties related to the process of installation and deployment. Therefore, it was decided to develop and take care of distribution internally.

It was opted for preconfigured Virtual Machine distribution for its portability and possibility to replicate Chemotargets software environment. Thanks to up-to-date development, cross-platform support and its open-source base, the choice feell on Oracle VM VirtualBox [6]. Its portability has a back side, as it could cause difficulties for copy protection.

The Linux distribution of a choice of Chemotargets is Debian [7,8], therefore it served as basis for the virtual appliance.

Code protection

Among other alternatives that are able to compile and freeze a python code, the method developed by Anthony Tuininga - *cx_Freeze* [9] - was used.

cx_Freeze consists of scripts and modules that allow to convert Python scripts, including their library dependencies, into executables. It also embeds the corresponding interpreter to the compiled code.

The code compilation transforms human-readable and interpreted open-source script to machine-readable closed-source bytecode.

Execution control and license management

Another requirement for commercial software is license-based execution control. For the implementation of this task, various elements were developed.

System-dependent unique identification

In order to identify every virtual appliance that is distributed, the python module based on dmidecode source code was employed [10,11]. Dmidecode dumps a computer's DMI (SMBIOS) table which contains a description of the system's hardware components, as well as other useful pieces of information such as serial numbers and BIOS revision.

When the virtual appliance is imported to the VirtualBox manager, it is identified by unique ID that can be accessed by the python-dmidecode module.

Encryption module

The protection of license authenticity can be achieved using data encryption of a license key file provided to clients. The python module that was implemented is called Simple Crypt [12], that wraps the The PyCrypto Toolkit [13] for more simple use. The cipher employed is AES256.

The same module is used to cypher inter-script messages detailed below.

License control module

The control of the license authenticity and its limits is aggregated in python-scripted and compiled daemon that starts during the system boot and is consulted by the main programs every execution.

The license management module receives encrypted information from the license key file, provided by Chemotargets, where the following information is encoded:

1. verification code corresponding to the unique system ID
2. time limit: number of days passed since the activation
3. molecules limit: a total number of molecules that can be processed
4. parallel execution limit: number of concurrent instances of CT-link
5. CPU limit: maximum number of cores that can be used

Furthermore, the license management module is also responsible for logging the usage and halt the CT-link execution if any limit is exhausted. The log is encrypted and the module controls any possible attempt of its external manipulation.

Summarizing, the license control module plays a central role in execution management of CT-link commands by communicating with them via secure protocol and logging the license limits exhaustion.

Inter-script communication module

The most widely used IPC (Inter-Process Communication) mechanism in GNU/Linux is D-Bus [14]. It was specifically designed for efficient and easy-to-implement messaging between programs running on the same system. Applications that use D-Bus typically connect to a bus daemon, which forwards messages between them.

In case of CT-link distribution, the license management module acts as a D-Bus daemon that listens to the queries from CT-link scripts. The answers are permissions or denials, depending on the current limits. All the messages that are sent between CT-link modules are encrypted so that they can not be falsified.

Results

The CT-link[CMD] is the command line version of the distribution and serves as backend server for the CT-link[GUI] frontend, that is developed by another PhD student at Chemotargets.

The CT-link software demo was requested by more than 20 companies and research groups worldwide. Moreover, 3 commercial (2 of them by major pharmaceutical companies) and 1 academic license were purchased during the first year.

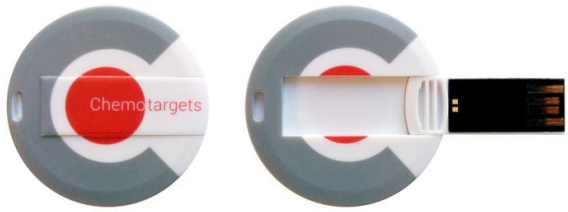

Figure 1
USB flash memory containing CT-Link trial distribution

References

1. Sanner MF. Python: a programming language for software integration and development. J Mol Graph Model. 1999;17: 57–61.
2. Python Programming Language [Internet]. Available: www.python.org
3. Kernighan BW, Ritchie D. The C Programming Language. Prentice-Hall; 1978.
4. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. J Cheminform. 2011;3: 33.
5. Landrum G. RDKit: Open-source cheminformatics [Internet]. Q1 2015 [cited Sep 2015]. Available: http://www.rdkit.org
6. Oracle VM. VirtualBox [Internet]. [cited 2015]. Available: www.virtualbox.org
7. Debian Project. Debian GNU/Linux 7 [Internet]. [cited 2015]. Available: http://www.debian.org
8. Aoki O. Debian Reference [Internet]. Dec 2014. Available: https://www.debian.org/doc/manuals/debian-reference/index.en.html
9. Tuininga A. cx_Freeze (4.3.4) documentation [Internet]. [cited 2015]. Available: http://cx-freeze.readthedocs.org/en/stable
10. Talebi N. Python-dmidecode package [Internet]. [cited 2015]. Available: https://packages.debian.org/wheezy/python-dmidecode
11. Cox A, Delvare J, Arapov A. Dmidecode [Internet]. [cited 2015]. Available: http://www.nongnu.org/dmidecode/
12. Cooke A. Simple Crypt [Internet]. [cited 2015]. Available: pypi.python.org/pypi/simple-crypt

13. Litzenberger D. PyCrypto - The Python Cryptography Toolkit [Internet]. [cited 2015]. Available: https://www.dlitz.net/software/pycrypto/

14. freedesktop.org. D-Bus [Internet]. [cited 2015]. Available: http://www.freedesktop.org/wiki/Software/dbus/

Part IV: Discussions

In silico approach for target deconvolution

The drug target identification process is similar to the reverse-engineering approach. The researcher has the information on the effect/perturbation that input signals (molecules) produce on the system (cells, organs or entire organisms) and she/he needs to understand the underlying mechanisms that will lead to the resulting complex state.

In the context of the current Thesis, I contributed to identify targets that are involved in antiproliferative effects of drugs on colon cancer HCT116 cell lines. Differential *in silico* pharmacological profiling led to the discovery of 115 targets that were annotated to selective tumor cytotoxic compounds and not to selective cytotoxic compounds on normal cells. A significant part of them were retrospectively validated to be oncogenes.

It was also confirmed that polypharmacological drugs are more efficient when it comes to perturbing a resistant and adaptive, yet robust, biological system, such as a tumor cell.

Biologically-diverse compound selection

The design of a diverse chemical library is usually based on chemical properties of the molecules and the selection is done by structural comparison of molecules.

In this Thesis, pharmacological fingerprints were used as criteria for selection of diverse compounds for a chemical library designed to be the core screening collection of the EU-OPENSCREEN initiative. Both known and predicted affinity data for more than a million molecules resulted in almost 400 000 unique target profiles, which served for

measuring the biological similarity between compounds and for selecting the most diverse ones, while offering maximum target coverage. When compared to other selection approaches, our bioactivity-based selection provided less similar overlapping molecules and demonstrated the additive value of accounting for target diversity when designing a chemical library.

Further applications of bioactivity prediction

The link describing the network of interactions between molecules and their targets is able to contribute to many additional fields. Some of the contributions in which I exploited and analyzed these data are summarized below.

Off-target related toxicity prediction was implemented at Sanofi using the Chemotargets' *in silico* target profiling software CT-link. A promising 81.1% predictive accuracy was obtained when affinity predictions for proposed targets were compared to internal Sanofi *in vitro* data. Moreover, it was possible to expand public-data models with internal data of the company to enhance both chemical and target coverage.

Systems biology approach to safety prediction

During my time at Chemotargets, I assisted in the preparation of the review of the existing *in silico* fragment-based predictive safety and collaborated in the development of the advanced biological system framework that takes into account molecular similarity, pharmacological data and related biological pathways.

Information about the expression of drug targets expression across different organs and tissues provides another layer of complexity to

our arsenal of predictive safety approaches. The anatomical projection of molecules, composed of expression scores of the targets extracted from their pharmacological profiles, gives insight on which organs are related both to primary targets and secondary pharmacology. The anatomical projection was used for analyzing a number of sets with known toxicity (namely, cardiotoxicity, hepatotoxicity, and torsadogenicity) and it was able to identify the organs enriched within the corresponding molecules in each organotoxic set.

The new “Anatomical” entity was added to the biological system framework, connecting organs with targets and, through targets, molecules with projected anatomical entities.

Commercial software distribution

The preconfigured virtual appliance with Chemotargets software (CT-link) became an easy approach to test the methodology at remote research sites, both in academia as well as the pharmaceutical industry. As a result, it led to various successful deals to acquire commercial licenses of CT-link.

During my PhD stay at Chemotargets, I understood the value of applied software. No matter how good is your methodology, if it is not developed with a commercial mind-setting or its distribution is complicated, it will never become a commercial product and be used at pharmaceutical companies in “real” drug discovery projects.

Future research directions

As can be observed in Figure 3 of the Introduction section, there are a number of biological system entities still left to connect to the global

framework. Some of them are already in the process of being implemented, others are at the very early stages of development. Among them, the anatomical information generated in this Thesis is now an integral part of the the safety prediction module in CT-link.

Another more distant, but potentially more important direction of the research, is to model not only the static state of the system, but also to take into account the dynamic changes introduced by small molecule perturbations. For example, besides annotation of proteins to their specific organ expressions levels, we can extract information on how drugs modify these expressions. In fact, there is an increasingly amount of GWAS data perturbed by drugs that is likely to have a key impact in future developments in this area. Or, to put another example, we could predict not only inhibition/activation of protein targets, but also how these changes in the activity influence the biological pathways where these proteins are involved and produce a knock-on effect that ultimately leads to serious adverse reactions. Last but not least, integration of all data on protein-protein interactions will also offer another viewpoint to the way small molecules perturb biological systems.

As for commercial software, there is also room for improvement. Code optimizations, alternative python compiler and reduction of a number of dependencies will allow the distribution of native software, leaving aside the virtualization layer. That could be interesting for large-scale enterprise servers that are limited now by the total number of 32 CPUs that VirtualBox can handle.

Part V: Conclusions

The main contributions of this Thesis can be summarized in the following points:

1. The framework of an interconnected biological system was defined. In the context of the current Thesis “Drugs”, “Targets”, “Safety Terms” and “Organs” were considered as systemic entities and interactions between them were compiled, integrated, modeled, predicted, stored, and managed
2. *In silico* prediction of drug-protein interactions was successfully applied to the identification of targets likely to be relevant in colon cancer.
3. Predicted bioactivity data of more than a million of compounds was used to select 40 000 of the most biodiverse molecules, extending chemical-based approaches to collaborative chemical library design within the remit of the EU-OPENSCREEN European project.
4. A novel computational method to predictive drug safety was developed and applied to real case scenarios within an industrial setting.
5. A novel computational approach to projecting the anatomical profile of drugs by integrating its pharmacological profile and the corresponding gene expression profiles was developed. The approach was applied as a proxy indicator of potential in vivo organ toxicities, namely, cardiotoxicity and hepatotoxicity

6. A new platform-independent distribution package for easy deployment of Chemotargets' software CT-link was developed. That has allowed the company to reach markets previously inaccessible and proving the true value of an industrial PhD program.

Appendix

Supporting table S1

The *in vitro* pharmacological profile of drugs as a proxy indicator of potential *in vivo* organ toxicities

AS.00.00.00.00.00.00.00	Anatomical System
AS.01.00.00.00.00.00.00	unclassifiable
AS.02.00.00.00.00.00.00	nervous system
AS.02.01.00.00.00.00.00	peripheral nervous system
AS.02.01.01.00.00.00.00	sciatic nerve
AS.02.01.02.00.00.00.00	sympathetic chain
AS.02.01.02.01.00.00.00	cervical sympathetic chain
AS.02.01.02.01.01.00.00	inferior cervical ganglion
AS.02.01.02.01.02.00.00	middle cervical ganglion
AS.02.01.02.01.03.00.00	superior cervical ganglion
AS.02.01.03.00.00.00.00	ganglion
AS.02.01.03.01.00.00.00	sympathetic ganglion
AS.02.01.03.01.01.00.00	celiac ganglion
AS.02.01.03.01.02.00.00	cervicothoracic ganglion
AS.02.01.03.02.00.00.00	nodose ganglion
AS.02.01.03.03.00.00.00	cranial sensory ganglion
AS.02.01.03.03.01.00.00	trigeminal ganglion
AS.02.01.03.04.00.00.00	spinal ganglion
AS.02.01.04.00.00.00.00	peripheral nerve
AS.02.01.05.00.00.00.00	olfactory apparatus
AS.02.01.06.00.00.00.00	auditory apparatus
AS.02.01.06.01.00.00.00	internal ear
AS.02.01.06.01.01.00.00	membranous labyrinth
AS.02.01.06.01.01.01.00	saccule
AS.02.01.06.01.01.02.00	utricle
AS.02.01.06.01.02.00.00	osseous labyrinth
AS.02.01.06.01.02.01.00	cochlear duct
AS.02.01.06.01.02.01.01	stria vascularis
AS.02.01.06.01.02.02.00	vestibule
AS.02.01.06.01.02.03.00	semicircular canal
AS.02.01.06.01.02.04.00	cochlea
AS.02.01.06.01.02.04.01	spiral organ of Corti
AS.02.01.06.02.00.00.00	auditory ossicle
AS.02.01.06.03.00.00.00	auditory tube
AS.02.01.06.04.00.00.00	tympanum
AS.02.01.06.05.00.00.00	middle ear
AS.02.01.06.06.00.00.00	external ear
AS.02.01.06.06.01.00.00	external acoustic meatus
AS.02.01.06.06.02.00.00	auricle
AS.02.01.07.00.00.00.00	eye
AS.02.01.07.01.00.00.00	extraocular muscle

AS.02.01.07.02.00.00.00	trabecular meshwork
AS.02.01.07.03.00.00.00	optic nerve
AS.02.01.07.04.00.00.00	retina
AS.02.01.07.04.01.00.00	fovea centralis
AS.02.01.07.04.02.00.00	macula lutea
AS.02.01.07.05.00.00.00	choroid
AS.02.01.07.06.00.00.00	ciliary body
AS.02.01.07.07.00.00.00	iris
AS.02.01.07.08.00.00.00	vitreous humor
AS.02.01.07.09.00.00.00	lens
AS.02.01.07.09.01.00.00	lens cortex
AS.02.01.07.10.00.00.00	sclera
AS.02.01.07.11.00.00.00	cornea
AS.02.01.07.12.00.00.00	conjunctiva
AS.02.01.07.13.00.00.00	lacrimal gland
AS.02.01.07.14.00.00.00	eyelid
AS.02.01.07.15.00.00.00	globe
AS.02.02.00.00.00.00.00	central nervous system
AS.02.02.01.00.00.00.00	spinal cord
AS.02.02.01.01.00.00.00	lumbosacral nucleus
AS.02.02.01.02.00.00.00	accessory nucleus
AS.02.02.01.03.00.00.00	ventromedial column
AS.02.02.01.04.00.00.00	ventrolateral column
AS.02.02.01.05.00.00.00	phrenic nucleus
AS.02.02.01.06.00.00.00	dorsolateral column
AS.02.02.01.07.00.00.00	dorsomedial column
AS.02.02.01.08.00.00.00	retrodorsolateral column
AS.02.02.01.09.00.00.00	ventral column
AS.02.02.01.10.00.00.00	sacral parasympathetic nucleus
AS.02.02.01.11.00.00.00	intermediomedial column
AS.02.02.01.12.00.00.00	intermediolateral column
AS.02.02.01.13.00.00.00	lateral column
AS.02.02.01.14.00.00.00	visceral column
AS.02.02.01.15.00.00.00	nucleus thoracicus
AS.02.02.01.16.00.00.00	nucleus proprius
AS.02.02.01.17.00.00.00	substantia gelatinosa
AS.02.02.01.18.00.00.00	dorsal column
AS.02.02.02.00.00.00.00	brain
AS.02.02.02.01.00.00.00	meninges
AS.02.02.02.01.01.00.00	pia mater
AS.02.02.02.01.02.00.00	arachnoid
AS.02.02.02.01.03.00.00	dura mater
AS.02.02.02.02.00.00.00	cerebrospinal fluid
AS.02.02.02.03.00.00.00	ventricular system
AS.02.02.02.03.01.00.00	choroid plexus
AS.02.02.02.03.02.00.00	fourth ventricle
AS.02.02.02.03.03.00.00	cerebral aqueduct
AS.02.02.02.03.04.00.00	third ventricle
AS.02.02.02.03.05.00.00	lateral ventricle

AS.02.02.02.04.00.00.00	tract
AS.02.02.02.04.01.00.00	corpus callosum
AS.02.02.02.05.00.00.00	cerebellum
AS.02.02.02.05.01.00.00	cerebellum nuclei
AS.02.02.02.05.01.01.00	nucleus fastigii
AS.02.02.02.05.01.02.00	nucleus globosus
AS.02.02.02.05.01.03.00	nucleus emboliformis
AS.02.02.02.05.01.04.00	dentate nucleus
AS.02.02.02.05.02.00.00	cerebellar cortex
AS.02.02.02.05.02.01.00	vermis
AS.02.02.02.05.02.02.00	flocculonodular lobe
AS.02.02.02.05.02.03.00	middle lobe of the cerebellum
AS.02.02.02.05.02.04.00	anterior lobe of the cerebellum
AS.02.02.02.05.03.00.00	cerebellum peduncles
AS.02.02.02.06.00.00.00	brain stem
AS.02.02.02.06.01.00.00	locus coeruleus
AS.02.02.02.06.02.00.00	medulla oblongata
AS.02.02.02.06.02.01.00	nucleus ambiguus
AS.02.02.02.06.02.02.00	nucleus intercalatus
AS.02.02.02.06.02.03.00	arcuate nuclei
AS.02.02.02.06.02.04.00	nucleus parasolitarius
AS.02.02.02.06.02.05.00	nucleus of the tractus solitarius
AS.02.02.02.06.02.06.00	dorsal vagal nucleus
AS.02.02.02.06.02.07.00	nucleus of the hypoglossal nerve
AS.02.02.02.06.02.08.00	accessory cuneate nucleus
AS.02.02.02.06.02.09.00	nucleus of the spinal tract of the trigeminal
AS.02.02.02.06.02.10.00	spinal nucleus of the accessory nerve
AS.02.02.02.06.02.11.00	supraspinal nucleus
AS.02.02.02.06.02.12.00	nucleus cuneatus
AS.02.02.02.06.02.13.00	nucleus gracilis
AS.02.02.02.06.02.14.00	olivary nuclei
AS.02.02.02.06.02.14.01	dorsal accessory
AS.02.02.02.06.02.14.02	medial accessory
AS.02.02.02.06.02.14.03	inferior olivary nuclei
AS.02.02.02.06.03.00.00	pons
AS.02.02.02.06.03.01.00	trigeminal nucleus
AS.02.02.02.06.03.01.01	principal sensory
AS.02.02.02.06.03.01.02	motor
AS.02.02.02.06.03.01.03	nucleus of the spinal tract
AS.02.02.02.06.03.02.00	salivatory nuclei
AS.02.02.02.06.03.02.01	inferior salivatory nuclei
AS.02.02.02.06.03.02.02	superior salivatory nuclei
AS.02.02.02.06.03.03.00	facial nucleus
AS.02.02.02.06.03.04.00	abducent nucleus
AS.02.02.02.06.03.05.00	nucleus of the lateral lemniscus
AS.02.02.02.06.03.06.00	trapezoid nucleus
AS.02.02.02.06.03.07.00	superior olivary nucleus
AS.02.02.02.06.03.08.00	cochlear nuclei
AS.02.02.02.06.03.08.01	ventral

AS.02.02.02.06.03.08.02	dorsal
AS.02.02.02.06.03.09.00	vestibular nuclei
AS.02.02.02.06.03.09.01	interstitial
AS.02.02.02.06.03.09.02	inferior vestibular nuclei
AS.02.02.02.06.03.09.03	superior vestibular nuclei
AS.02.02.02.06.03.09.04	lateral
AS.02.02.02.06.03.09.05	medial
AS.02.02.02.06.04.00.00	midbrain
AS.02.02.02.06.04.01.00	subcommissural organ
AS.02.02.02.06.04.02.00	mesencephalic trigeminal nucleus
AS.02.02.02.06.04.03.00	trochlear nucleus
AS.02.02.02.06.04.04.00	oculomotor nucleus
AS.02.02.02.06.04.05.00	periaqueductal grey matter
AS.02.02.02.06.04.06.00	red nucleus
AS.02.02.02.06.04.07.00	substantia nigra
AS.02.02.02.06.04.08.00	colliculi
AS.02.02.02.06.04.08.01	inferior colliculi
AS.02.02.02.06.04.08.02	superior colliculi
AS.02.02.02.06.04.09.00	crus cerebri
AS.02.02.02.06.04.10.00	ventral tegmentum
AS.02.02.02.07.00.00.00	diencephalon
AS.02.02.02.07.01.00.00	hypothalamus
AS.02.02.02.07.01.01.00	lateral tuberal nucleus
AS.02.02.02.07.01.02.00	lateral mamillary nucleus
AS.02.02.02.07.01.03.00	medial mamillary nucleus
AS.02.02.02.07.01.04.00	tuberomamillary nucleus
AS.02.02.02.07.01.05.00	premamillary nucleus
AS.02.02.02.07.01.06.00	posterior nucleus
AS.02.02.02.07.01.07.00	lateral nucleus
AS.02.02.02.07.01.08.00	ventromedial nucleus
AS.02.02.02.07.01.09.00	dorsomedial nucleus
AS.02.02.02.07.01.10.00	anterior nucleus
AS.02.02.02.07.01.11.00	infundibular nucleus
AS.02.02.02.07.01.12.00	paraventricular nucleus
AS.02.02.02.07.01.13.00	suprachiasmatic nucleus
AS.02.02.02.07.01.14.00	supraoptic nucleus
AS.02.02.02.07.01.15.00	preoptic nucleus
AS.02.02.02.07.02.00.00	subthalamus
AS.02.02.02.07.02.01.00	subthalamic nucleus
AS.02.02.02.07.03.00.00	epithalamus
AS.02.02.02.07.03.01.00	habenular nucleus
AS.02.02.02.07.03.02.00	pineal body
AS.02.02.02.07.04.00.00	metathalamus
AS.02.02.02.07.04.01.00	lateral geniculate nucleus
AS.02.02.02.07.04.02.00	medial geniculate nucleus
AS.02.02.02.07.05.00.00	thalamus
AS.02.02.02.07.05.01.00	limiting thalamic nucleus
AS.02.02.02.07.05.02.00	centromedian thalamic nucleus
AS.02.02.02.07.05.03.00	reticular thalamic nucleus

AS.02.02.02.07.05.04.00	pulvinar
AS.02.02.02.07.05.05.00	lateral posterior thalamic nucleus
AS.02.02.02.07.05.06.00	lateral dorsal thalamic nucleus
AS.02.02.02.07.05.07.00	lateral thalamic nuclei
AS.02.02.02.07.05.08.00	ventral posterior thalamic nucleus
AS.02.02.02.07.05.09.00	ventral intermediate thalamic nucleus
AS.02.02.02.07.05.10.00	ventral anterior thalamic nucleus
AS.02.02.02.07.05.11.00	ventral thalamic nuclei
AS.02.02.02.07.05.12.00	central lateral thalamic nucleus
AS.02.02.02.07.05.13.00	paracentral thalamic nucleus
AS.02.02.02.07.05.14.00	submedial thalamic nucleus
AS.02.02.02.07.05.15.00	parafascicular thalamic nucleus
AS.02.02.02.07.05.16.00	medial dorsal thalamic nucleus
AS.02.02.02.07.05.17.00	medial thalamic nuclei
AS.02.02.02.07.05.18.00	anterior ventral thalamic nucleus
AS.02.02.02.07.05.19.00	anterior medial thalamic nucleus
AS.02.02.02.07.05.20.00	anterior dorsal thalamic nucleus
AS.02.02.02.07.05.21.00	anterior thalamic nuclei
AS.02.02.02.08.00.00.00	cerebrum
AS.02.02.02.08.01.00.00	basal nuclei
AS.02.02.02.08.01.01.00	amygdala
AS.02.02.02.08.01.01.01	cortical amygdaloid nucleus
AS.02.02.02.08.01.01.02	medial amygdaloid nucleus
AS.02.02.02.08.01.01.03	central amygdaloid nucleus
AS.02.02.02.08.01.01.04	amygdaloid nucleus
AS.02.02.02.08.01.02.00	globus pallidus
AS.02.02.02.08.01.02.01	globus pallidus medial
AS.02.02.02.08.01.02.02	globus pallidus lateral
AS.02.02.02.08.01.03.00	putamen
AS.02.02.02.08.01.04.00	lentiform nucleus
AS.02.02.02.08.01.05.00	caudate nucleus
AS.02.02.02.08.01.06.00	corpus striatum
AS.02.02.02.08.01.07.00	claustrum
AS.02.02.02.08.01.08.00	accumbens
AS.02.02.02.08.02.00.00	cerebral cortex
AS.02.02.02.08.02.01.00	parahippocampal gyrus
AS.02.02.02.08.02.02.00	cingulate gyrus
AS.02.02.02.08.02.03.00	hippocampus
AS.02.02.02.08.02.04.00	secondary olfactory cortex
AS.02.02.02.08.02.05.00	primary olfactory cortex
AS.02.02.02.08.02.06.00	olfactory tubercle
AS.02.02.02.08.02.07.00	medial olfactory stria
AS.02.02.02.08.02.08.00	lateral olfactory stria
AS.02.02.02.08.02.09.00	anterior olfactory nucleus
AS.02.02.02.08.02.10.00	olfactory bulb
AS.02.02.02.08.02.11.00	insula
AS.02.02.02.08.02.12.00	visual
AS.02.02.02.08.02.13.00	occipital lobe
AS.02.02.02.08.02.13.01	visual cortex

AS.02.02.02.08.02.14.00	temporal lobe
AS.02.02.02.08.02.15.00	parietal lobe
AS.02.02.02.08.02.16.00	frontal lobe
AS.02.02.02.08.02.16.01	prefrontal cortex
AS.03.00.00.00.00.00.00	dermal system
AS.03.01.00.00.00.00.00	appendages
AS.03.01.01.00.00.00.00	hair root
AS.03.01.02.00.00.00.00	sebaceous gland
AS.03.01.03.00.00.00.00	sweat gland
AS.03.01.04.00.00.00.00	nail
AS.03.01.05.00.00.00.00	nail bed
AS.03.01.06.00.00.00.00	hair
AS.03.01.07.00.00.00.00	hair follicle
AS.03.02.00.00.00.00.00	skin
AS.03.02.01.00.00.00.00	dermis
AS.03.02.01.01.00.00.00	dermal papilla
AS.03.02.02.00.00.00.00	epidermis
AS.03.02.02.01.00.00.00	stratum corneum
AS.03.03.00.00.00.00.00	adipose tissue
AS.04.00.00.00.00.00.00	musculoskeletal system
AS.04.01.00.00.00.00.00	fascia
AS.04.02.00.00.00.00.00	tendon
AS.04.03.00.00.00.00.00	muscle
AS.04.03.01.00.00.00.00	smooth muscle
AS.04.03.01.01.00.00.00	iris sphincter muscle
AS.04.03.01.02.00.00.00	vascular smooth muscle
AS.04.03.02.00.00.00.00	skeletal muscle
AS.04.03.02.01.00.00.00	fast-twitch skeletal muscle
AS.04.03.02.02.00.00.00	slow-twitch skeletal muscle
AS.04.03.02.03.00.00.00	longissimus muscle
AS.04.03.02.03.01.00.00	longissimus dorsi muscle
AS.04.04.00.00.00.00.00	joint
AS.04.04.01.00.00.00.00	articular cartilage
AS.04.04.02.00.00.00.00	meniscus
AS.04.04.03.00.00.00.00	ligament
AS.04.04.04.00.00.00.00	synovium
AS.04.04.04.01.00.00.00	synovial fluid
AS.04.05.00.00.00.00.00	cartilage
AS.04.06.00.00.00.00.00	bone
AS.04.06.01.00.00.00.00	cancellous bone
AS.04.06.02.00.00.00.00	growth plate
AS.04.06.03.00.00.00.00	epiphysis
AS.04.06.04.00.00.00.00	femur
AS.04.06.05.00.00.00.00	condyle
AS.04.06.06.00.00.00.00	tibia
AS.05.00.00.00.00.00.00	endocrine system
AS.05.01.00.00.00.00.00	thymus
AS.05.02.00.00.00.00.00	adrenal gland
AS.05.02.01.00.00.00.00	adrenal medulla

AS.05.02.02.00.00.00.00	adrenal cortex
AS.05.02.02.01.00.00.00	zona glomerulosa
AS.05.03.00.00.00.00.00	parathyroid
AS.05.04.00.00.00.00.00	thyroid
AS.05.05.00.00.00.00.00	pituitary gland
AS.05.05.01.00.00.00.00	posterior pituitary
AS.05.05.02.00.00.00.00	anterior pituitary
AS.05.05.02.01.00.00.00	pars tuberalis
AS.05.06.00.00.00.00.00	pineal gland
AS.05.07.00.00.00.00.00	endocrine pancreas
AS.05.07.01.00.00.00.00	islet of Langerhans
AS.06.00.00.00.00.00.00	urogenital system
AS.06.01.00.00.00.00.00	reproductive system
AS.06.01.01.00.00.00.00	female reproductive system
AS.06.01.01.01.00.00.00	colostrum
AS.06.01.01.02.00.00.00	blastocyst
AS.06.01.01.02.01.00.00	blastocyst inner cell mass
AS.06.01.01.02.02.00.00	trophoblast
AS.06.01.01.03.00.00.00	umbilical cord
AS.06.01.01.03.01.00.00	umbilical artery
AS.06.01.01.03.02.00.00	umbilical vein
AS.06.01.01.04.00.00.00	milk
AS.06.01.01.05.00.00.00	breast
AS.06.01.01.05.01.00.00	mammary duct
AS.06.01.01.05.02.00.00	mammary gland
AS.06.01.01.05.03.00.00	nipple
AS.06.01.01.06.00.00.00	amniotic fluid
AS.06.01.01.07.00.00.00	amnion
AS.06.01.01.08.00.00.00	chorioamniotic membrane
AS.06.01.01.09.00.00.00	placenta
AS.06.01.01.10.00.00.00	vulva
AS.06.01.01.11.00.00.00	vagina
AS.06.01.01.12.00.00.00	uterus
AS.06.01.01.12.01.00.00	corpus
AS.06.01.01.12.02.00.00	cornu
AS.06.01.01.12.03.00.00	myometrium
AS.06.01.01.12.04.00.00	endometrium
AS.06.01.01.12.05.00.00	uterine cervix
AS.06.01.01.12.05.01.00	ectocervix
AS.06.01.01.13.00.00.00	uterine tube
AS.06.01.01.14.00.00.00	ovary
AS.06.01.01.14.01.00.00	corpus luteum
AS.06.01.01.14.02.00.00	ovarian follicle
AS.06.01.01.14.02.01.00	ovarian follicular fluid
AS.06.01.01.14.02.02.00	ovarian follicle membrane
AS.06.01.02.00.00.00.00	male reproductive system
AS.06.01.02.01.00.00.00	semen
AS.06.01.02.01.01.00.00	seminal plasma
AS.06.01.02.02.00.00.00	penis

AS.06.01.02.02.01.00.00	corpus spongiosum
AS.06.01.02.02.02.00.00	corpus cavernosum
AS.06.01.02.02.03.00.00	foreskin
AS.06.01.02.02.04.00.00	glans
AS.06.01.02.03.00.00.00	vas deferens
AS.06.01.02.03.01.00.00	seminal vesicle
AS.06.01.02.04.00.00.00	prostate
AS.06.01.02.05.00.00.00	epididymis
AS.06.01.02.06.00.00.00	testis
AS.06.01.02.06.01.00.00	seminiferous tubule
AS.06.01.02.06.02.00.00	testis germ cells
AS.06.01.02.06.03.00.00	testis interstitial cells
AS.06.01.02.06.04.00.00	testis leydig cells
AS.06.02.00.00.00.00.00	urinary system
AS.06.02.01.00.00.00.00	urothelium
AS.06.02.02.00.00.00.00	urethra
AS.06.02.03.00.00.00.00	bladder
AS.06.02.04.00.00.00.00	ureter
AS.06.02.05.00.00.00.00	kidney
AS.06.02.05.01.00.00.00	renal medulla
AS.06.02.05.01.01.00.00	outer medulla of kidney
AS.06.02.05.01.02.00.00	inner medulla of kidney
AS.06.02.05.02.00.00.00	renal cortex
AS.06.02.05.03.00.00.00	nephron
AS.06.02.05.03.01.00.00	renal tubule
AS.06.02.05.03.01.01.00	renal proximal tubule
AS.06.02.05.03.01.01.01	renal proximal convoluted tubule
AS.06.02.05.03.01.02.00	renal collecting duct
AS.06.02.05.03.01.03.00	loop of Henle
AS.06.02.05.03.01.04.00	renal distal convoluted tubule
AS.06.02.05.03.02.00.00	renal corpuscle
AS.06.02.05.03.02.01.00	glomerus
AS.07.00.00.00.00.00.00	alimentary system
AS.07.01.00.00.00.00.00	saliva
AS.07.02.00.00.00.00.00	gastrointestinal tract
AS.07.02.01.00.00.00.00	esophagogastric junction
AS.07.02.02.00.00.00.00	intestine
AS.07.02.02.01.00.00.00	large intestine
AS.07.02.02.01.01.00.00	cecum
AS.07.02.02.01.02.00.00	anus
AS.07.02.02.01.03.00.00	colorectal
AS.07.02.02.01.03.01.00	rectum
AS.07.02.02.01.03.02.00	colon
AS.07.02.02.01.03.02.01	ascending colon
AS.07.02.02.01.03.02.02	proximal colon
AS.07.02.02.01.03.02.03	distal colon
AS.07.02.02.01.03.02.04	appendix
AS.07.02.02.02.00.00.00	vermiform appendix
AS.07.02.02.03.00.00.00	small intestine

AS.07.02.02.03.01.00.00	intestinal crypt
AS.07.02.02.03.02.00.00	intestinal villus
AS.07.02.02.03.03.00.00	ileum
AS.07.02.02.03.04.00.00	jejunum
AS.07.02.02.03.05.00.00	duodenum
AS.07.02.03.00.00.00.00	stomach
AS.07.02.03.01.00.00.00	gastric fundus
AS.07.02.03.01.01.00.00	gastric fundic mucosa
AS.07.02.03.02.00.00.00	gastric antrum
AS.07.02.03.03.00.00.00	antral mucosa
AS.07.02.03.04.00.00.00	pyloric antrum
AS.07.02.03.05.00.00.00	stomach cardia
AS.07.02.04.00.00.00.00	esophagus
AS.07.03.00.00.00.00.00	pancreas
AS.07.03.01.00.00.00.00	pancreatic duct
AS.07.03.02.00.00.00.00	exocrine pancreas
AS.07.03.02.01.00.00.00	pancreatic acinus
AS.07.03.02.01.01.00.00	pancreatic juice
AS.07.04.00.00.00.00.00	liver and biliary system
AS.07.04.01.00.00.00.00	bile duct
AS.07.04.02.00.00.00.00	gall bladder
AS.07.04.03.00.00.00.00	liver
AS.07.04.03.01.00.00.00	bile
AS.07.05.00.00.00.00.00	peritoneum
AS.07.05.01.00.00.00.00	peritoneal cavity
AS.07.06.00.00.00.00.00	omentum
AS.07.06.01.00.00.00.00	lesser omentum
AS.07.06.02.00.00.00.00	greater omentum
AS.07.07.00.00.00.00.00	mesentery
AS.07.08.00.00.00.00.00	pharynx
AS.07.08.01.00.00.00.00	hypopharynx
AS.07.08.02.00.00.00.00	oropharynx
AS.07.08.03.00.00.00.00	nasopharynx
AS.07.09.00.00.00.00.00	oral cavity
AS.07.09.01.00.00.00.00	lip
AS.07.09.02.00.00.00.00	periodontium
AS.07.09.02.01.00.00.00	periodontal ligament
AS.07.09.03.00.00.00.00	tooth bud
AS.07.09.04.00.00.00.00	jaw
AS.07.09.04.01.00.00.00	mandible
AS.07.09.05.00.00.00.00	palate
AS.07.09.06.00.00.00.00	gum
AS.07.09.07.00.00.00.00	tooth
AS.07.09.07.01.00.00.00	incisor
AS.07.09.07.02.00.00.00	enamel organ
AS.07.09.07.03.00.00.00	dental pulp
AS.07.09.07.04.00.00.00	tooth enamel
AS.07.09.07.05.00.00.00	molar
AS.07.09.08.00.00.00.00	salivary gland

AS.07.09.08.01.00.00.00	sublingual gland
AS.07.09.08.02.00.00.00	submandibular gland
AS.07.09.08.03.00.00.00	parotid gland
AS.07.09.09.00.00.00.00	tongue
AS.07.09.09.01.00.00.00	circumvallate papilla
AS.07.09.09.02.00.00.00	Ebner's gland
AS.07.09.09.03.00.00.00	taste bud
AS.07.09.09.03.01.00.00	foliate papilla
AS.08.00.00.00.00.00.00	lymphoreticular system
AS.08.01.00.00.00.00.00	Peyer's patch
AS.08.02.00.00.00.00.00	spleen
AS.08.03.00.00.00.00.00	tonsil
AS.08.03.01.00.00.00.00	palatine tonsil
AS.08.03.02.00.00.00.00	pharyngeal tonsil
AS.08.03.03.00.00.00.00	lingual tonsil
AS.08.04.00.00.00.00.00	lymph node
AS.08.04.01.00.00.00.00	germinal center
AS.08.05.00.00.00.00.00	lymph
AS.09.00.00.00.00.00.00	hematological system
AS.09.01.00.00.00.00.00	blood
AS.09.01.01.00.00.00.00	blood plasma
AS.09.01.02.00.00.00.00	blood serum
AS.09.02.00.00.00.00.00	bone marrow
AS.10.00.00.00.00.00.00	respiratory system
AS.10.01.00.00.00.00.00	diaphragm
AS.10.02.00.00.00.00.00	pleura
AS.10.03.00.00.00.00.00	lung
AS.10.04.00.00.00.00.00	bronchus
AS.10.04.01.00.00.00.00	bronchial epithelial cells
AS.10.04.02.00.00.00.00	bronchiole
AS.10.04.02.01.00.00.00	alveolus
AS.10.05.00.00.00.00.00	trachea
AS.10.06.00.00.00.00.00	larynx
AS.10.07.00.00.00.00.00	sinus
AS.10.08.00.00.00.00.00	nose
AS.10.08.01.00.00.00.00	nasal cavity
AS.10.08.02.00.00.00.00	nasal vestibule
AS.11.00.00.00.00.00.00	cardiovascular system
AS.11.01.00.00.00.00.00	blood vessel
AS.11.01.01.00.00.00.00	capillary
AS.11.01.02.00.00.00.00	vein
AS.11.01.02.01.00.00.00	saphenous vein
AS.11.01.02.02.00.00.00	portal vein
AS.11.01.02.03.00.00.00	jugular vein
AS.11.01.02.04.00.00.00	venous adventitia
AS.11.01.02.05.00.00.00	venous media
AS.11.01.02.06.00.00.00	venous intima
AS.11.01.02.07.00.00.00	vena cava
AS.11.01.03.00.00.00.00	artery

AS.11.01.03.01.00.00.00	femoral artery
AS.11.01.03.02.00.00.00	pulmonary artery
AS.11.01.03.03.00.00.00	coronary artery
AS.11.01.03.04.00.00.00	carotid artery
AS.11.01.03.05.00.00.00	arterial adventitia
AS.11.01.03.06.00.00.00	arterial media
AS.11.01.03.07.00.00.00	arterial intima
AS.11.01.03.08.00.00.00	aorta
AS.11.02.00.00.00.00.00	heart
AS.11.02.01.00.00.00.00	cardiac conducting system
AS.11.02.02.00.00.00.00	cardiac valve
AS.11.02.03.00.00.00.00	pericardium
AS.11.02.04.00.00.00.00	myocardium
AS.11.02.04.01.00.00.00	cardiac myocytes
AS.11.02.05.00.00.00.00	endocardium
AS.11.02.06.00.00.00.00	atrioventricular node
AS.11.02.07.00.00.00.00	heart ventricle
AS.11.02.07.01.00.00.00	left ventricle
AS.11.02.08.00.00.00.00	heart atrium
AS.11.02.08.01.00.00.00	right atrium
AS.12.00.00.00.00.00.00	immune system
AS.12.01.00.00.00.00.00	lymphoid cells
AS.12.01.01.00.00.00.00	T cells
AS.12.01.02.00.00.00.00	B cells
AS.12.01.03.00.00.00.00	NK cells
AS.12.02.00.00.00.00.00	myeloid cells
AS.12.02.01.00.00.00.00	mast cell
AS.12.02.02.00.00.00.00	basophil
AS.12.02.03.00.00.00.00	eosinophil
AS.12.03.00.00.00.00.00	monocytes
AS.12.03.01.00.00.00.00	macrophages
AS.12.03.02.00.00.00.00	dentritic cells