Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biology

Kristine Lippestad

Automation of Drug Target Profiling for Logical Modelling

Master's thesis in Biotechnology
Supervisor: Åsmund Flobak
Co-supervisor: Martin Kuiper and Eirini Tsirvouli
May 2023

NTNU
Norwegian University of
Science and Technology

Kristine Lippestad

Automation of Drug Target Profiling for Logical Modelling

Master's thesis in Biotechnology
Supervisor: Åsmund Flobak
Co-supervisor: Martin Kuiper and Eirini Tsirvouli
May 2023

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biology

NTNU
Norwegian University of
Science and Technology

Abstract

Cancer is a heterogeneous group of diseases that poses a major global health problem. Today, a great deal of resources are allocated to acquire cancer control. The significant progress in cancer research has, among various advancements, resulted in targeted therapies that represent a groundbreaking approach to cancer treatment. Targeted drugs are developed to selectively modify molecules that constitute the cell circuitry for deciding on cancer growth, progression, and metastasis, in contrast to traditional chemotherapeutics that indiscriminately target the growth machinery of cells.

Despite the aim of selectivity, it is frequently observed that additional target molecules, referred to as off-targets, are affected. *In silico* modelling is a valuable approach to predict how cancer as a system responds to perturbations. However, off-target effects are rarely considered in simulations of targeted therapies, which is a bottleneck to achieving accurate predictions of systems responses.

To understand the impact of off-target effects, *in silico* modelling was conducted to identify how the predicted drug effect was changed. A relational database was implemented to efficiently handle such off-targets, storing information on drugs, targets and the prior knowledge available on drug-target interactions. The database content was made accessible by a tool developed to carry out customised queries, facilitating automated drug target profiling accounting for off-target effects. The new module was applied to generate target profiles for 17 drugs. The interpretation of these presented that several drugs inhibited multiple molecules in addition to their primary targets. Further investigation of the mechanism of inhibition suggested that low selectivity resulted from highly conserved target regions. Additionally, it was observed that drugs with high off-target activity frequently showed low selectivity towards a kinase subfamily. The target profiles were applied in logical modelling to simulate drug effects. Prediction of cell viability implied that off-targets tend to enhance the drug responses when the modified entities were members of distinct signalling pathways. Simulations with a more robust network implied that modification of multiple random targets has a low impact, suggesting that the topology of a target node should be emphasised in drug selection to promote the effect. Through simulations with the DrugLogics pipeline, it was illustrated that accounting for off-target effects in the perturbation data improved the software's capability to predict drug synergies. Enhanced prediction performance was confirmed from raised values for the area under the curve (AUC) obtained from the plotted receiver operating characteristic (ROC) and precision-recall (PR) curves. The applicability of enhanced drug target profiling was again demonstrated by applying the tool to characterise the target profile of the *Imatinib* candidate target therapy. The characterisation enabled off-label use based on the overlap between mutations identified in the patient's tumour and the off-target profile of the drug.

The results of this project emphasise the importance of including off-target effects when carrying out *in silico* simulation of drug response for enhanced prediction performance. This improvement is enabled by the developed module that performs automated drug target profiling, accounting for off-target effects.

Sammendrag

Kreft er en heterogen sykdomsgruppe som utgjør en betydelig global helseutfordring. I dag legges det inn store ressurser for å oppnå kreftkontroll. Betydelig framgang innen kreftforskning har resultert i flere nye og lovende behandlingsmetoder for kreft, der i blant målrettede legemidler. Målrettede legemidler er utviklet til å selektivt påvirke molekyler som utgjør det molekylære beslutningsapparatet for cellevekst, progresjon og metastase av kreftsykdommen, i motsetning til tradisjonell cellegift som angriper cellenes vekstmaskineri.

På tross av intensjonen om selektivitet, er det ofte observert at målmolekyler som ikke var tilsiktet, kalt "off-targets", påvirkes av målrettede legemidler. *In silico* modellering er en nyttig metode for å predikere hvordan kreft, som et system, responderer på perturbasjoner. "Off-target-effekter" blir sjeldent tatt i betrakting i simulering av målrettede legemidler, og dette representerer en begrensning for å oppnå presise predikasjoner av systemets respons.

For å forstå innvirkningen av "off-target-effekter", ble det gjennomført *in silico* modellering med mål om å identifisere hvordan responsen på legemidlene endres. For å tilrettelegge for effektiv håndtering av slike "off-target-effekter" ble det utviklet en relasjonsdatabase, som lagrer informasjon om medisiner, deres målmolekyler og eksisterende kunnskap om interaksjonen mellom disse. Databaseinnholdet ble tilgjengeliggjort gjennom et verktøy laget for å gjennomføre skreddersydde spørringer for å tilrettelegge for automatisk generering av målprofiler for medisiner, som inkluderer "off-target-effekter". Den nye modulen ble anvendt for å finne målprofilene til 17 medisiner, og analysen av disse viste at en rekke inhiberte flere molekyler i tillegg til primær målmolekylene. Videre undersøkelser av inhibisjonsmekanismen antydet at lav selektivitet kan skyldes høy konservering av målregionen. Det ble og observert at medisiner med høy "off-target-aktivitet" ofte viste lav selektivitet mot en enkelt kinasegruppe. Målprofilene ble brukt i logisk modellering for å simulere effekten av medisinene. Predikering av cellevekst viste at "off-target-effekter" ofte forsterker responsen når de påvirkede molekylene tilhører ulike signalspor. Simuleringer utført på et mer robust nettverk indikerte at inhibering av mange tilfeldige målmolekyler hadde liten effekt, som antyder at topologien til målnoden bør tas i betraktning for å oppnå best mulig effekt. Gjennom simuleringer med DrugLogic pipelinen ble det illustrert at inkludering av "off-target-effekter" i perturbasjonsdataen fremmer programvarens evne til å predikere medisinsynergier. En forbedret predikasjonsytelse ble bekreftet fra økte verdier for areal under kurven (AUC) som ble kalkulert for plottede receiver operating characteristic (ROC) and precision-recall (PR) kurver. Nytten av forbedrete målprofiler ble videre demonstrert da verktøyet ble anvendt for å karakterisere målprofilen til det målrettede legemiddelet *Imatinib*, for å muliggjøre off-label bruk basert på hvordan mutasjonene, identifisert i en pasients svulst, sammenfalt med "off-target-profilen" til medisinen.

Resultatene i dette prosjektet understreker viktigheten av å inkludere "off-target-effekter" ved *in silico* simuleringer av respons på medisiner for å oppnå bedre predikasjonsytelse. Forbedret ytelse er lagt til rette for gjennom utvikling av en ny modul som utfører automatisk generering av målprofiler som inkluderer "off-target-effekter".

Preface and Acknowlegdement

This master's thesis marks the end of my Master of Science in Biotechnology at the Norwegian University of Science and Technology (NTNU). The thesis was conducted at the Department of Biology with the DrugLogics initiative, under the supervision of Åsmund Flobak, from January 2022 to May 2023.

I want to express my gratitude to my main supervisor, Åsmund Flobak, and my co-supervisors, Martin Kuiper and Eirini Tsirvouli, for your much appreciated guidance and encouragement during the work with my thesis. Furthermore, I would like to thank you for the possibility of developing the project in a manner that has enabled me to use a great deal of the knowledge I have acquired during my degree, including molecular biology, biochemistry, programming and database development. Last but not least, allowing me to gain experience in front-end development, which I have highly enjoyed. I would also like to show my appreciation towards my fellow students in the DrugLogics initiative for inspiring discussions and motivating words.

Finally, I'm very grateful to my family and friends for always supporting me and believing in me throughout my five years of study. I want to express a special thanks to my flatmates, who have been my family away from home, and my dear friends from my study program, who have brought much joy to my everyday life at NTNU.

Trondheim, May 2023

Kristine Lippestad

List of Abbreviations

$\boldsymbol{IC}_{50}$ half maximal inhibitory concentration.

$\boldsymbol{K}_d$ dissociation constant.

$\boldsymbol{K}_i$ inhibition constant.

AGS adenocarcinoma cancer cell line.

AUC area under the curve.

CASCADE CAncer Signaling CAusality DatabasE.

CLO Cell Line Ontology.

DBMS database management system.

DT drug-target.

DTI drug-target interaction.

EFO Experimental Factor Ontology.

ER Entity-Relationship.

FN False Negative.

FP False Positive.

FPR False Positive Rate.

GDSC Genomics of Drug Sensitivity in Cancer.

GUI Graphical User Interface.

HSA Highest Single Agent model.

InChi Key The International Chemical Identifier.

mAbs Monoclonal antibodies.

PKN prior knowledge networks.

PR precision-recall.

ROC receiver operating characteristic.

SMILES Simplified Molecular Input Line Entry Specification.

TN True Negative.

TP True Positive.

TPR True Positive Rate.

UniProtKB Universal Protein Resources KnowledgeBase.

Table of Contents

1 Introduction

The global cancer burden poses a significant challenge to human well-being, causing one in six deaths worldwide in 2020 [1]. In 134 of 184 countries in the world, cancer is currently the first or second most common cause of premature death, and approximately one in five individuals will develop this disease during their life [2, 1]. The figures continue to increase, and unfortunately, the burden is predicted to be expanded by 47% from 2020 to 2040 [3]. The root causes of the expansion are multifaceted, with raised life expectancy and prominent alteration in lifestyle due to improved economy playing prominent roles [3].

Alongside the growing cancer prevalence, cancer research is currently undergoing substantial progress. Extensive research is resulting in enhanced knowledge of cancer prevention and treatment. Primary prevention is facilitated by identifying major risk factors to avoid the onset of cancer—for instance, use of tobacco, alcohol, dietary risks and air pollution. Furthermore, secondary prevention is improved, mainly through screening programs [4]. Many countries have now introduced systematic screening programs such as mammographic screening for early detection of breast cancer and pap smears for cervical cancer [3]. Screening programs support an early and precise cancer diagnosis, allowing for timely and effective treatment [4]. Nevertheless, the development of tertiary prevention is advanced, with new and better therapeutics frequently put on the market [4]. The treatment strategy is slowly progressing from a "one-size-fits-all" approach to precision medicine led by new technologies such as immunotherapy and targeted therapy [5].

1.1 The Hallmarks of Cancer

Cancer is a heterogeneous group of diseases that share a set of functional capabilities, ultimately resulting in uncontrolled tissue growth. The functional capabilities were summarised as the six hallmarks of cancer by Hanahan et al. in 2000 [6]. These capabilities emerge from an accumulation of mutations, genomic rearrangements and epigenetic changes that activate oncogenes or inactivate tumour suppressor genes. As a result, the activity of signal transduction pathways that regulate cell fate, genome maintenance, and cell survival is modified [7]. The modifications make cancer cells capable of survival, proliferation and spread, defying normal cell control [6].

The first and second hallmarks are *sustaining proliferative signalling* and *evading growth suppressors*, which commonly initiate the process of tumourigenesis. The first capability causes the cells to induce and sustain growth-promoting signals, thereby disrupting homeostasis. The second hallmark lets the cancer cells deregulate programs that prevent cell growth and is often brought about by inactivation of tumour suppressor genes. *Resisting cell death* enables the cancer cells to avoid programmed cell death by apoptosis when exposed to stress and DNA damage, which would normally eliminate otherwise healthy cells. The next trait, *enabling replicative immortality*, acquires the cancer cells to bypass senescence and crisis. Senescence is a nonproliferative yet viable state. Crisis is a phase where most cells in a population die. These barriers of unlimited proliferation are usually avoided by the elongation or stabilisation of telomeres, which protect the end of the chromosomes. *Inducing angiogenesis* is the capability of stimulating the growth of new blood vessels. These blood vessels provide the tumour with nutrients and oxygen and dispose of metabolic wastes and carbon dioxide, enhancing tumourigenesis. The final capability, *activating invasion and metastasis*, enables the cancer cell to move from local tissue and vessels to new sites, transported through the circulation, and establish secondary tumours [6].

The six hallmarks were later extended with additional capabilities by Hanahan et al. in 2011 and 2022. The new capabilities include four emerging and four enabling hallmarks [8, 9]. The first emerging hallmark is *deregulating cellular energetics*, representing modification or reprogramming

of the cellular metabolism in cancer cells, a capability that enhances continuous cell growth and proliferation. *Avoiding immune destruction* acquires cancer cells to bypass immunological destruction, specifically by T and B lymphocytes, macrophages and natural killer cells [8]. Furthermore, *unlocking phenotypic plasticity* leads to avoidance of the terminal differentiation state to allow for cancer pathogenesis. The last emerging hallmark is the capability of *senescent cells*, which usually undergo proliferative arrest that cannot be reversed, to resume the proliferative state. The result is promoted tumour progression and malignant development [9].

The first enabling characteristic, *genomic instability*, results from genetic modifications that support tumourigenesis. Secondly, *avoiding immune destruction*, alter the functionality of immune cells, originally aiming to oppose infections and repair lesions, to enhance several of the other hallmark capabilities [8]. The hallmarks can also be enabled by *nonmutational epigenetic reprogramming* that cause modified gene expression. The final enabling characteristic is *polymorphic microbiomes*, representing the ecosystem of bacteria and fungi that can protect cancer progression and affect the response to therapy [9].

Altogether, these hallmarks represent the underlying foundation of capabilities, leading to complex phenotypes and genotypes across cancer diseases.

1.2 Cancer Therapy

Cancer has traditionally been treated by surgery, radiation therapy and cytotoxic chemotherapy. For disease that cannot be controlled surgically or radiotherapeutically, systemic therapies are used. Although systemic treatments like chemotherapies are important approaches to reducing rapid can cell growth, they tend to act unselectively. Consequently, vulnerable healthy growing cells are damaged. Therefore, the doses are kept restricted to protect the healthy cells, and a combination of numerous cytotoxic agents is required. The result is severe side effects that are challenging to endure [7, 10].

1.2.1 Molecular Targeted Therapy

A substantial shift in the development of cancer therapy is now taking place. Cytotoxic chemotherapy that directly targets mechanisms of cell growth, such as DNA synthesis, is gradually being substituted by molecularly targeted therapies. These therapies are designed to act selectively on molecular targets to counteract cancer cell survival, proliferation, and metastasis [11]. Inhibition of the hallmarks of cancer is enabled by modifying dysregulated proteins or pathways that are identified to drive tumourigenesis [12]. By addressing specific alterations of the individual patient, cancer therapy is enhanced from the "one-size-fits-all" approach applied in cytotoxic chemotherapy to precision medicine.

The targeted therapy approach is traditionally founded upon the 'one gene, one drug, one disease' paradigm. The paradigm is established from the agreement between genetic reductionism, which assumes that knowledge about genes is adequate to understand every feature of human behaviour, and novel molecular biology methods allowing for precise validation of single 'disease-causing' genes [13]. Successful drug design of a ligand that selectively modifies the disease-causing gene would lead to the fulfilment of the concept of 'magic bullets' originally posed by Ehrlich to combat infectious diseases. The philosophy is based on the idea of only eliminating the disease source without harming the body itself [14].

The two main types of targeted therapy are Monoclonal antibodies (mAbs) and small molecule inhibitors. mAbs target extracellular proteins and prevent interaction between a receptor and

the associated ligand [15]. The agents prevent direct interaction by binding to an antigen, cell receptor, or membrane-bound protein. Alternatively, or additionally, the mechanism can activate a subsequent response of the body's defence mechanism often referred to as 'antibody dependent cellular cytotoxicity'. The body's immune system can then, for instance, act by immune system detection of cancer cell-specific antigens that recruit effector cells and cause cytotoxic or phagocytic responses [16].

The second group is small molecule inhibitors. These drugs are characterised by low molecular weight (< 900 Da) and the ability to infiltrate cells to target a protein of interest [15]. Small molecule inhibitors most often function by inactivating kinases to regulate signalling pathways which are altered during tumourigenesis [15]. Alternative mechanism of action are targeting proteasomes, cyclin-dependent kinases (CDKs), or poly ADP-ribose polymerase (PARP) inhibitors that interfere with DNA damage repair which ultimately triggers apoptosis. Small molecule inhibitors, is proven successful in many patients [17]. An example of proof of concept in the clinic is using the tyrosine kinase inhibitor imatinib, an ABL inhibitor, on chronic myeloid leukaemia patients whose driver mutation is a BCR-ABL translocation [18].

Although molecular target therapy has shown noteworthy clinical success, some crucial drawbacks must be considered. For the therapy to be successful, a patient's tumour must express the specific biomarker that the agent modifies, and tumour growth must be critically dependent on this. Furthermore, despite the presence of the biomarker, drug resistance is a challenge in numerous patients. It is commonly observed that the drug initially has a good effect, but only a partial response is obtained before the development of therapy resistance. Resistance can be developed due to the complexity of oncogenic pathway crosstalk caused by multiple mutations or a lack of specificity where the agent cannot target the mutation of interest [15]. Resistance might also arise from characteristics such as drug efflux, alteration of the drug target, DNA damage repair, inactivation of cell death, epithelial-mesenchymal transition, and the cell heterogeneity in the tumour microenvironment [19]. Furthermore, a drug can lead to a strong response but simultaneously show side effects. Side effects are commonly caused by either on- or off-target effects. On-target effects are caused by an unintended mechanism of action on the target of interest, which can cause dysregulation of the molecule's function in normal cells. Off-target effects are caused by a lack of specificity, causing the drug to inhibit additional unintended targets [20].

1.2.2 Multi-Targeted Drugs

Bioactivity studies such as *in vitro* biochemical assays, high-throughput screening, and genomics have, in retrospect, shown that numerous small molecules used in the clinic are indeed multi-targeted [21]. Polypharmacological properties are acquired when a drug interacts with multiple macromolecular targets [22].

Off-target effect are often attributed to the binding site the drug is designed to target being a conserved region across multiple proteins. The most common example is the adenosin triphosphate (ATP)-binding pocket of kinases. Kinases are signalling biomolecules commonly deregulated in cancer, which have become an important drug target class. All kinases share the same enzymatic activity of transmitting the terminal phosphate of ATP to a new protein. The transfer is facilitated by binding the ATP substrate to the ATP binding pocket, which is functionally and structurally conserved across the complete set of protein kinases, referred to as the kinome [23]. Numerous inhibitors are designed to target this binding pocket and block ATP from binding, and cross-reactivity for additional kinases is often observed [21]. A well-known example of this is the earlier mentioned drug, *Imatinib*. Inhibitors are also observed to be conformationally adaptable toward structurally diverse conformational states of various kinases. This property is seen for both the

drug *Dasatinib* and *Erlotinib* [24].

Off-target effects can cause severe side effects, such as dysregulation of kinases which are essential for critical functions in the heart [25]. On the other hand, such effects can show potential for drug repurposing. In addition, multiple studies have shown that multi-targeted drugs can give an enhanced clinical effect and argue for the end of the single targeted therapy approach replaced with multi-targeted drugs [13, 26, 27]. Cancer phenotypes are often observed to be robust due to alternative compensatory signalling routes [28]. Moreoved, tumourigenesis is often driven by the modification of multiple pathways. Consequently, several proteins might be required to be inhibited to bring about an effect. Targeting multiple proteins could therefore be a promising strategy in cancer therapy [13].

1.2.3 Drug Synergies

Combining drugs is another example of an upcoming approach to treating cancer, by targeting multiple essential features or weaknesses. This method aims to identify a combination of drugs that show synergistic drug actions. A drug synergy is defined by the combination of drugs resulting in an enhanced effect compared to the sum of the effect of the individual drugs [29, 30]. Several synergies have already been proven to be successful. For instance, the combination of MEK and BRAF inhibitors [31], along with cyclin-dependent kinase 4/6 inhibitors and anti-estrogen therapies [32]. Synergies are often advantageous as the development of drug resistance can be postponed. In addition, therapeutic effects can be maintained despite a significant cut in the dosage of each drug, thereby reducing undesired side effects [29, 30].

1.3 Drug-Target Interactions

The ensemble of targets a drug interacts with is referred to as the target profile [33]. Mapping of drug-target interaction (DTI) is essential to predict the drug effect correctly and to acquire insight into a drug's mechanism of action and level of selectivity. This knowledge enables successful drug development, prediction of drug effect and cancer therapy selection.

Most drug targets are enzymes and small molecule drugs are typically developed to inhibit the enzyme activity. Inhibitors interact with their targets in three different manners. The most common mechanism is competitive inhibition, where the drug competes with the enzyme's substrate for the active site. The inhibitor then blocks the active site and stops the substrate from binding. As mentioned earlier, this is a common mechanism of action for kinase inhibitors, which occupy the ATP binding pocket. The second mechanism, uncompetitive inhibition, occurs through binding to a distinct site on the enzyme. Molecules that express this type of inhibition bind exclusively to the enzyme-substrate complex. The binding leads to the alteration of the enzyme conformation, reducing the binding affinity between the substrate and the active site. As a result, the catalysation rate is reduced. The final mechanism is noncompetitive inhibition. Equal to the uncompetitive mechanism, binding occurs in a region distinct from the active site and causes a change in enzyme conformation. However, the interaction occurs independent of substrate binding [34].

1.3.1 Binding Affinity

Drug target profiling is typically carried out using binding assays which measure the drug's binding affinity across a panel of numerous of potential targets. A binding affinity value report the strength of a DTI. Estimating to which degree an inhibitor binds to possible target enzymes is valuable for

evaluating if an interaction is adequate to obtain a therapeutic effect or for predicting if interactions with off-targets will cause undesired side effects [35].

The binding affinity strength between a drug-target (DT)-pair is generally expressed by the dissociation constant (K_d), the inhibition constant (K_i), or half maximal inhibitory concentration (IC_{50}) [36]. IC_{50} states the total concentration that decrease enzymatic activity by 50% [37]. K_d and K_i are dissociation constants and state the free concentration of the drug necessary for 50% enzyme saturation. While K_d is a general measurement of the binding affinity between a molecule and an enzyme, K_i is specified for an inhibitor and an enzyme. Targeted therapies are typically small molecules designed to inhibit altered molecules causing uncontrolled growth and survival [10]. K_i is therefore often used. Low values for the lot indicate high binding affinity [36].

The K_d is equivalent to the K_i and only depends upon the ligand (drug) and enzyme (target) concentration. In contrast, IC_{50} also depends on the substrate concentration and the experimental conditions [35]. Assay conditions should therefore be as similar as possible for IC_{50} values from different laboratories to agree. The relationship between measurements depends on the reaction mechanism and the type of inhibition. In the case of non-competitive inhibitors, IC_{50} equals K_d and K_i. In the case of competitive inhibition kinetics, IC_{50} is expected to be higher. The drug then competes with the substrate. Hence, the necessary drug concentration to decrease the enzymatic activity by 50% depends on the substrate concentration and how strongly it binds the enzyme. However, the three values become close to equal if the substrate concentration is low [37].

1.4 Systems Biology

Systems biology is founded on the paradigm that a system as a whole is greater than the sum of its parts and cannot fully be understood by analysing components or even pathways in isolation [38]. The belief is, therefore, that the structure and dynamics of a cell or organism should be studied rather than the characteristics of the single components alone to comprehend the complexity of living systems [39]. The holistic paradigm was introduced as an objection to the reductionist philosophy of René Descartes. Descartes stated that a complex system could be fully understood by reducing it to manageable elements, analysing them separately, and then assembling the whole from the characteristics of the individual components. Likewise, Jacques Loeb's mechanistic biology philosophy was replaced by systems biology. Mechanistic biology is based on the thought of all biological behaviour being set, forced, and uniform across all individuals of one species, making them nothing more than complex machines [38].

A key concept within systems biology is that living systems have a hierarchical organisation, a philosophy introduced by James Henry Woodger at the beginning of the 20th century. The concept states that, apart from the most basic level, every level of an organisation has a characteristic defined by the interactions across the elements in the lower levels. As an illustration, the organised cellular functions of regulatory proteins lead to a working cell cycle [38]. Ludwig von Bertalanffy and his general system theory later evolved the systems biology paradigm [38]. His theory describes how all systems are developed from interlinked components. Systems can therefore be described by a network where entites are represented as nodes and the interactions between them are represented by edges. The shared characteristic of interlinked components can result in similarities in the complex structure and control design of the network. This feature is demonstrated by stable system structures, which frequently are characterised by highly connected nodes, referred to as hubs, and edges [38]. Numerous biological systems show this behaviour, having a small set of proteins that are part of multiple pathways, forming hubs, while the majority only are included in one or two.

The field of systems biology has advanced significantly over the past decades, and is fuelled in part by the large data amounts associated with genomic characterisation [40]. The use of high-throughput sequencing approaches has allowed the development of network models to simulate complex biological systems, thereby improving the understanding of these systems [40]. These approaches are shown to be promising in multiple applications, for instance, to enhance the understanding of normal and perturbed disease conditions.

This area of application is called systems medicine. The approach utilise network models for the integration of molecular patient data. By interpreting the interactions within these systems, critical entities in the disease can be identified, facilitating the discovery of diagnostic and prognostic biomarkers [41].

1.4.1 Cancer as a Systems Biology Disease

Cancer is an excellent example of a disease that can be understood from a systems biology perspective. Cancer arises from alterations in genes involved in crucial signal transduction pathways that regulate the functional state of cells. These signal transduction pathways form a network of cellular circuits, which are fundamental lines of intracellular communication. Intracellular communication occurs through integrating and transmitting intra- and extracellular signals via these pathways, which alter gene expression, protein activity and cellular behaviour. Thereby allowing for transition between cellular states, such as cycle progression, apoptosis, angiogenesis, and infiltration, in response to the signal [7, 42].

The signalling pathways are interconnected at multiple hierarchical layers. Cross-talk can occur directly by, for example, phosphorylation and indirectly by, for example, regulation of gene expression. Furthermore, one molecule can be included in multiple pathways. The system can exhibit even more complex behaviour as a result of regulatory motifs, which highly influence the signal response [42]. These motifs are frequently repeating patterns that can amplify, dampen, prolong or shorten the transmitted signals. Negative and positive feedback loops are typical examples of such motifs, where pathway components positively or negatively influence the activity of upstream components of the pathway. The result is a complex, self-regulating biological system [43].

Systems biology gives valuable insight into the complex network of molecular interactions which elucidate the mechanism of cancer. Development of computational models from big data sets enables researchers to identify genes and pathways that drive cancer, allowing for the discovery of new drug targets [42]. Moreover, the approach can be applied to predict drug effects.

1.4.2 Biological Networks

Existing knowledge and experimental observations describing a biological system can be incorporated in a prior knowledge networks (PKN). The structure and organisation of nodes and edges in a network can capture the pairwise interactions between molecules, including receptor-ligand binding, enzyme-substrate interactions, or protein-protein interactions, that lead to cellular behaviour. A network edge, can either be undirected, indicating a two-way and symmetrical relationship, or directed, where component A regulates component B, but not the other way around. This edge is said to be outgoing for node A and incoming for node B, and depicts information flow. Regulatory networks are commonly described by directed graphs, were edges represent either inhibiting or activating interactions, indicating if communication along the edges leads to down-regulation or up-regulation, respectively. The graphs allow tracking a path of coupled biochemical reactions from the input node to the output node [43].

A scale-free network characteristic tends to be observed in regulatory networks [43]. Scale-free networks display a wide set of mathematically derived characteristics, of which those important for the project have been the power law degree distribution. A degree distribution (P(k)) represents the probability that a random node has exactly k edges. The parameter is computed by dividing the number of nodes (N(k)) of each degree by the total number of nodes. A power law distribution is characterised by a few nodes having a high degree, i.e. many edges, forming hubs, while most edges have a low degree [44].

1.5 Boolean Modelling

After developing a prior knowledge network that captures existing knowledge and experimentally observations of the cell behaviour, the network can be converted to an executable computational model based on a mathematical framework to perform simulations. The model is then exposed to different conditions and is enabled to change over time. The simulation results can be analysed to make qualitative or quantitative predictions about the cell behaviour when the system is perturbed.

The properties of computational models developed for biological systems vary depending on the mathematical framework used to describe and analyse the system. The status of the system components can be represented by continuous values, holding any real number within a set interval or discrete values. Furthermore, the model can allow for randomness or noise in the data leading to less predictable behaviour. The model is then classified as stochastic. In contrast, a deterministic model performs prediction based exclusively on the initial conditions and the regulatory rules defined for each node. Boolean models represent the most straightforward version, being discrete, deterministic and parameter-free [45].

The Boolean modelling approach is founded on the proposition logic defined by George Boole. The paradigm applies the principle of bivalence, meaning a statement must be either true or false, and a third possibility or contradiction is not considered an option [46]. Unlike continuous models, this approach does not require precise kinetic parameters, which makes it suitable for modelling large-scale systems with numerous variables and interactions without extensive experimental data. As a result, Boolean modelling can be used to analyse complex systems and predict how their behaviour changes over time.

1.5.1 Building a Boolean Model

A regulatory network can be translated into a Boolean model. Boolean models are commonly used to illustrate regulatory networks due to their ability to accurately depict the network topology. A Boolean model is implemented by assigning each node in the directed graph with a Boolean variable indicating the local activity state. The activity state can either be True (1), indicating that the node is active, or False (0), indicating that the node is inactive. As a result, a network consisting of N nodes will have a total of 2^{N} potential states. Furthermore, each node is associated with logical rules. A rule defines how the activity state of a node is determined by the state of the neighbouring regulators. The formulas use the logical operators NOT, AND, OR to derive conditions where the activity state is True or False. For instance, node A can be defined to have the activity state True if either of its activating nodes B OR C is active AND NOT its inhibitor node D. In short, the activity state of a node will stay the same or be altered according to execution the of the logical equation. An update of the activity state of a regulating node generally triggers a corresponding alteration in the state of its connected nodes until a stable state is reached. As a result, the network undergoes a dynamic trajectory, which illustrates how an input signal can evolve the network state over time [45].

Increasingly more comprehensive and complex regulatory networks involved in regulating multiple biological processes are developed as a result of extensive data generation and integration efforts. To allow for the modelling of such complex networks, multi-valued variables can be introduced [47]. As a result, the activity state is not restricted to True or False but can take on various values for intermediate activity levels, resulting in a logical model. A regulatory node can then control different targets at distinct levels and have multiple effects on a single target [48]. This approach can be convenient for predicting different levels of pro- or antisurvival represented by the output nodes, for example.

1.5.2 Simulations with a Logical Model

After implementing the model, simulations can be performed to predict cell behaviour when exposed to a condition of interest. The simulation can be carried out using an asynchronous or synchronous updating policy. In the case of synchronous update, each process, presented by an edge, is implicitly assumed to happen at the same rate. Therefore, node states are modified with regular intervals of a set time step. For each iteration, the activity level of every node is updated according to their logic formula before a new iteration is carried out. For this reason, the order of updates does not influence the outcome. All simulations starting with the same initiating conditions will result in the same state after an equal number of steps. Hence, the system is entirely deterministic. The trajectory of the system will eventually converge toward an attractor. The attractor is either a single state, referred to as a steady state, a repeating pattern of states, referred to as a cycle or an ensemble of stable states the system randomly oscillates between. Because synchronous updates cannot account for the variation in rate of signal propagation in a biological system, asynchronous updating policy is more commonly applied [45].

Asynchronous updating accounts for variation in timescales by introducing distinct time delays for every update of an activity state and running the updates in a non-synchronous manner. Hence, the activity state of each node is updated independently, according to the logical rule, without waiting for the activity state of the other components to be updated. Different attractors are reached for each simulation, resulting in less predictable behaviour. Asynchronous modelling is commonly carried out by running numerous replicate simulations with identical initial conditions. The reachable states and the likelihood for each can then be identified. The approach enables the prediction of the system behaviour with a high degree of certainty.

A system can be exposed to various conditions to predict how different types of perturbation influence the system. An example is predicting a drug's or drug combinations' effect [29, 30].

1.5.3 The DrugLogics Pipeline

The DrugLogics (NTNU) pipeline is a powerful example, showing the value of logical modelling. The DrugLogics initiative has developed software for predicting drug synergies *in silico*, which may be followed by validation through experimental testing. The result is that lab experiments can be guided by computer simulations, which can limit the experimental search space that would otherwise have to be covered. The modelling pipeline comprises several software modules that function cooperatively to enable a highly automated computational framework for logic model assembly and simulations [29, 30, 49].

The software optimises a general logical model to represent a specific cell line. The optimisation process obtains an ensemble of models generated by a genetic algorithm. By iteratively altering the logical rules, the algorithm produces models that that increasingly fit with the data provided

in a training file to resemble the cell line of interest. Next, drug perturbations are introduced systematically to the ensemble of models. A drug perturbation is simulated by fixing the activity state of a target node to be 0 or 1, representing inhibition and activation respectively. The effect of the various drug combinations is predicted by evaluating the output nodes' computed activity state that indicates cell viability. Finally, the predicted synergies are compared against observed synergies validated by experimental data to evaluate the prediction performance [49].

A limitation of the software is that no module for adequate drug target profiling is provided. Studies performed with the pipeline have only introduced perturbations of the primary targets of each drug. Hence, potential off-target effects are not accounted for, which is a critical bottleneck to allow for enhanced prediction performance. To allow for improved predictions the software would benefit from a module with a database of DTI-data and an associate tool for extensive target profiling.

1.6 Introduction to Database Design and Implementation

A significant amount of data is generated through biological research, and the development of database applications is essential to store this data efficiently. With the help of databases, data can be organised into a structured format, making it accessible for all researchers through straightforward queries and analyses.

The database design process begins with a requirement and collection analysis. The data requirements of the intended group of end users are then defined. The result is described in a miniworld, the extract of the real world to be represented in the database. The description includes the relevant objects, the relationship between these and the rules that will apply [50]. An example can be a database storing information about genes with a unique id, a name and a DNA sequence. A gene can encode proteins with a unique id, name, and primary structure provided by the amino acid sequence.

After collecting and analysing the requirements, a conceptual schema is designed for the database. A high-level conceptual data model is then applied, for example, an Entity-Relationship (ER) model. The model represents every object with a physical existence by an entity [50]. Furthermore, sets of entities with shared properties are collected into entity types. In this example, *Gene* and *Protein*. The properties that describe an entity, such as an id and a name, are called attributes. The attributes derive their possible values from a data domain, such as a string or an integer. An entity type typically has a unique key attribute for each entity, which can be used for identification. Each entity type is represented by a rectangle box with the entity type name and a list of attributes in the ER model. Furthermore, the key attribute is specified by an underline [50].

The association between entity types are described by relationship types — for example, a gene *Encodes* a protein. Diamond-shaped boxes with the relationship name represent the relationship types. A straight line is drawn from this box to each participating entity type. The number of relationship instances an entity can be part of is restricted by a cardinality ratio [50]. For instance, a protein is encoded by one gene (1, 1), and a gene can encode for zero to multiple proteins due to alternative splicing (0, N). Hence, the cardinality ratio for this relationship type is 1:N (one-to-many). The cardinality ratio is indicated along the lines drawn from the relationship type to the entity types. Figure 1.1a shows the ER diagram created for this example.

The ER model can be enhanced to include concepts such as specialisation. An entity type is then a superclass specialised into a set of subclasses. The specialisation distinguishes among characteristics of the entities in the superclass. For example, instances of a superclass *Tissue* can belong to different tissue types. A specialisation is illustrated by a circle notation. The ER model

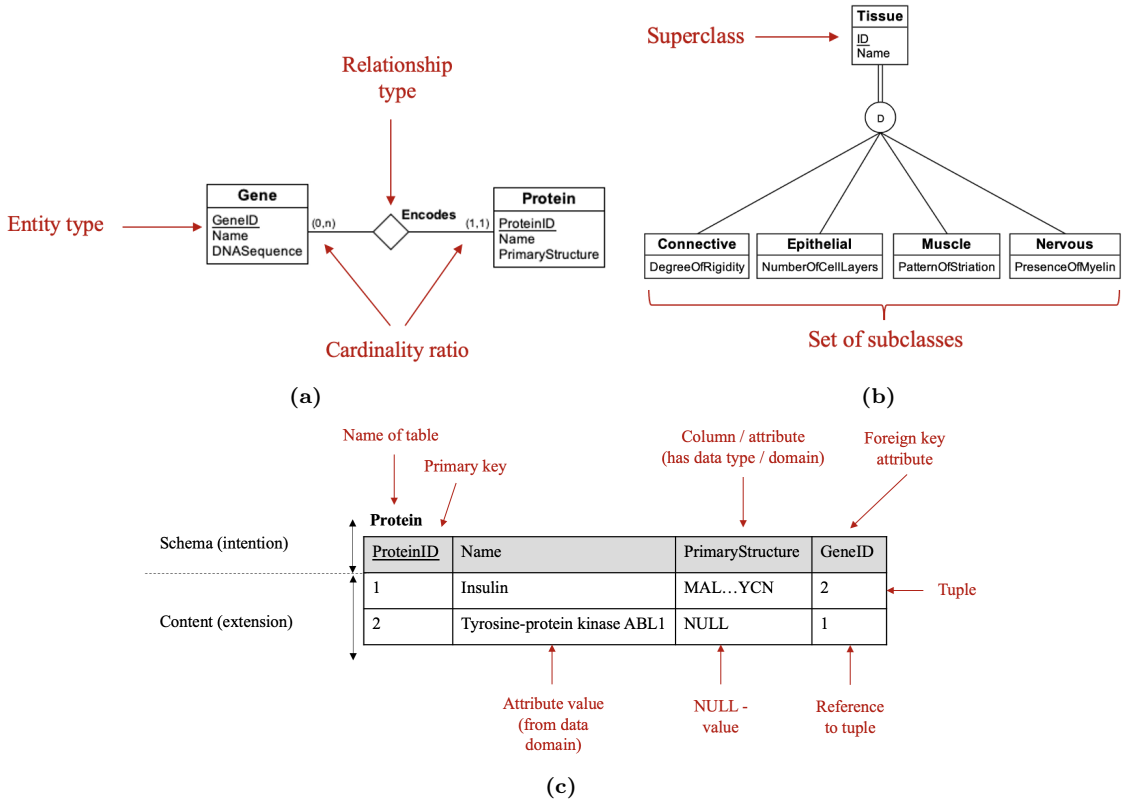
specifies whether the subclasses must be a disjoint set, indicated by a **d**, or if they can overlap, indicated by an **o**. An additional constraint defines if the specialisation is total. Hence, every entity must be part of at least one subclass or partial, where an entity does not need to be part of a subclass. This completeness constraint is indicated by a double or a single line, respectively [50]. Figure 1.1b illustrates the example of specialisation.

The next step in the database design process is called logical design. The ER schema is then translated into a relational database schema. The entity types are mapped to a relation (table). Each relation consists of a schema with the relation name, and a column with a defined data type for each attribute. In the relation extension, tuples for each entity are provided. A tuple consists of only valid values. If a value is unknown, it is assigned a NULL value [50]. Figure 1.1c shows the relation developed for the *Protein* relation.

An important task in this design step is to ensure entity and reference integrity. Entity integrity is achieved when each tuple in a relation is unique, which is ensured by a primary key constraint. The primary key represent a unique identifier for a tuple and cannot be NULL.

The relationship types are indicated by foreign keys that point to the primary key of a tuple in the referenced relation. When the cardinality ratio is 1:N, the foreign key is typically included in the relation, which can only be related to one tuple in the related relation. In the case of an M:N (many-to-many) cardinality ratio, a new relationship relation is created. The relation includes primary keys from each participating relationship, and the combination of these is the primary key of the relationship relation. In addition, they are independent foreign keys that reference each of the related relations. Reference integrity is achieved by only having foreign keys that point to an existing tuple or consist of NULL values [50].

In the final step, called physical design, a physical database model of a database management system (DBMS) is created from the database schema. Additional specifications are then defined for how the data should be stored and accessed. When the database design is completed, the database is loaded with actual data [50].



(c)

Figure 1.1: Central concepts within database design. a) Example of an Entity-Relationship (ER) schema diagram storing data for the entity types *Gene* and *Protein* associated by the relationship type *Encodes*. b) Example of an enhanced ER schema to illustrate specialisation. A set of subclasses {*Connective*, *Epithelial*, *Muscle*, *Nervous*} is a a specialisation of the superclass *Tissue*. c) Relation representing the *Protein* entity type with an associated schema and extension with tuples for each protein.

1.7 Project Objectives

The key motivation for this project was to study how model based predictions are changed when off-target effects are included in simulations of drug response. The project is founded upon the observation that targeted therapies tend to be multi-targeted. Knowledge of off-target effects is therefore anticipated to give better insight into how a drug operates, facilitating enhanced prediction of drug response. A new module of the DrugLogics pipeline is therefore required, to enable drug target profiling which accounts for off-target effects. Below is a brief description of the project's main objectives and the biological questions to be answered.

Part I - Development of a module connected to the DrugLogics pipeline to facilitate drug target profiling, accounting for off-target effects:

1. Construct a relational database providing data on anti-cancer drugs, their targets and the interaction between DT-pairs by
   (a) designing a relational database framework for efficient data storage
   (b) collecting data of interest from multiple freely available data resources
   (c) implementing programs that homogenise and organise the collected data before loading the data into the designed database
2. Develop a DrugProfiler tool associated with the database, aiming to
   (a) make the data accessible to users independent of their programming background
   (b) execute database queries according to user-specified conditions to generate customised drug target profiles
   (c) make the query result applicable for use in the DrugLogics pipeline
   (d) implement a Graphical User Interface (GUI) with text and figures allowing for visualisation and analysation of the generated target profiles

   Part II - Use the new module to generate drug target profiles and apply these in drug effect prediction:
3. Perform drug target profiling using the DrugProfiler tool to answer
   (a) How selective are the studied inhibitors?
   (b) Does the molecular mechanism of inhibitor activity impact drug selectivity?
   (c) How are the kinase targets within each profile distributed across the human kinome?
4. Perform *in silico* modelling of drug effect to answer
   (a) Does accounting of off-target effects alter simulated cell viability, as predicted by logical modelling?
   (b) Does the prediction performance of drug synergies improve when off-target effects are considered?
5. Use the DrugProfiler tool to investigate if a cancer patient could benefit from treatment with a off-label drug by comparing the generated target profile with the mutations identified in the patient's tumour.

2 Materials and Methods

2.1 Workflow Overview

As stated in the introduction, the main objective of this master's project was to investigate how model based predictions of drug response are altered when including off-target effects. For that, an efficient handling of such off-target effects is needed. This requirement is realised through a relation database that provides detailed information and context to drug-target interactions (DTI). The database development was initiated by a requirement and collection analysis to get an overview of the data needed to fulfil the needs of the end users. A database was then designed according to the analysis results. The required data were simultaneously collected from multiple knowledge databases and the literature. The database was implemented by structuring and homogenising the data before loading it into the designed framework.

After finalising the database, the DrugProfiler tool was developed to enable the automated generation and visualisation of drug target profiles. Code was implemented to conduct customised queries according to user-specified conditions to produce the profiles. Furthermore, a Graphical User Interface (GUI) providing figures and text were established from the output to allow for interactive exploration of the retrieved information.

The tool was applied to identify target profiles for 17 drugs for which the DrugLogics already had produced extensive results. The output was examined through a literature study on drugs' inhibition mechanisms. Furthermore, phylogenetic trees of the human kinome were annotated with the profile components to investigate the composition and distribution of drug targets.

The target profiles were used in logical modelling. Firstly, stochastic simulations were performed to examine how *in silico* prediction of drug effect was altered when accounting for off-target effects. Secondly, the DrugLogics pipeline was applied to investigate if the prediction of drug synergies could be improved by including off-target effects.

Finally, the DrugProfiler tool was applied to evaluate if a patient could benefit from treatment with a off-label drug. The identified target profile was examined according to the patient's molecular profile to see if the drug inhibited one or more of the proteins encoded by the mutated genes.

The project workflow is summarised in Figure 2.1.

2.2 Database Design

A database application was developed with a database and associated programs to perform database updates and queries. The objective was to create a database that stores drug-target interaction data, including off-target effects, to facilitate extensive drug target profiling. The profiles can, for instance, be applied in *in silico* simulations to improve predictions of cell behaviour when perturbed with single drugs or drug combinations. Furthermore, the data can facilitate cancer treatment selection according to a patient's molecular profile, allowing for precision medicine. Figure 2.2 shows an overview of the main steps in the database design.

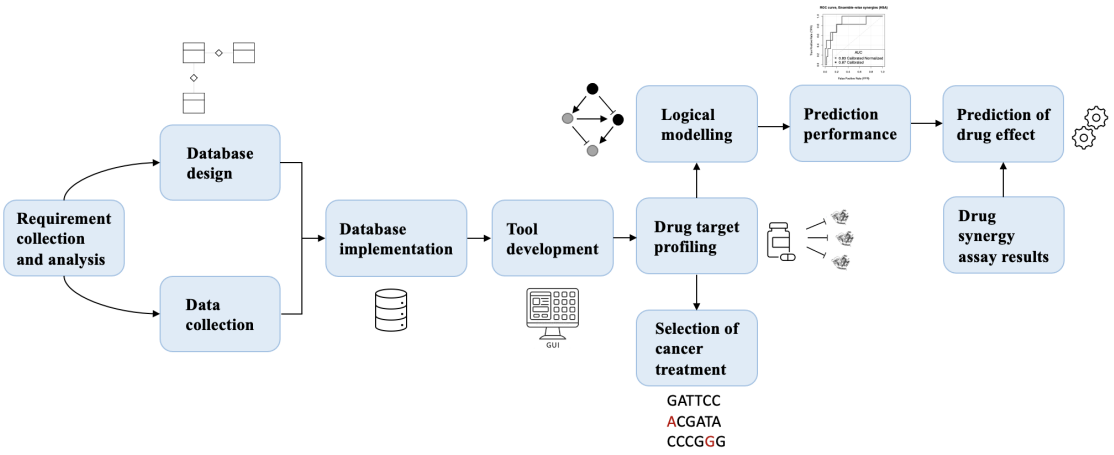


Figure 2.1: A brief summary of the stages involved in construction of a module to facilitate automated drug target profiling, and application of the module for prediction of drug effect and selection of cancer treatment. First, a database was designed meeting a set of user requirements and loaded with data collected from multiple data resources. Next, an associated tool was developed to allow for straightforward querying of the database to generate customised drug target profiles. The tool was then applied to identify and analyse target profiles for a set of drugs. The target profiles were used in logical modelling to predict drug effect, and it was evaluated if taking off-target effects into consideration improved the prediction performance by comparing the result with gold standards retrieved from drug synergy assays results. Finally, the tool was applied in the drug selection process for a cancer patient.

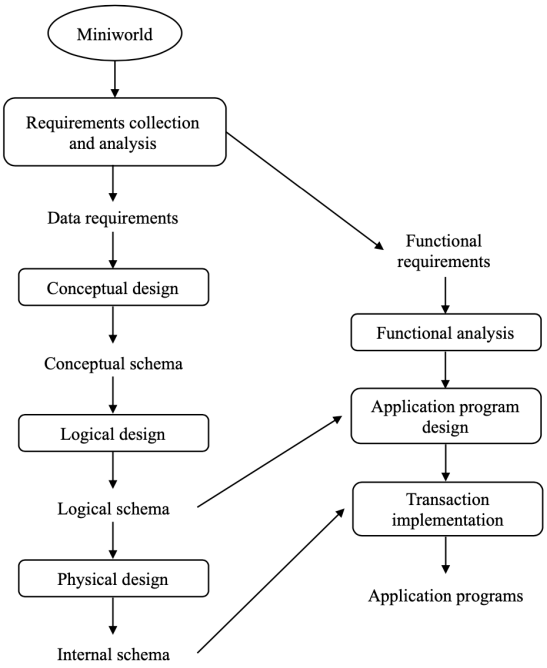


Figure 2.2: Flowchart showing the main phases of the design process to develop the drug-target interaction database.

2.2.1 Requirement Collection and Analysis

The initial phase of the database development process was requirement collection and analysis. The objective of this phase was to define the data requirements to fulfil the needs of the DrugLogics initiative which is expected to deliver the primary users.

The first step was to examine the results and data used by DrugLogics in previous studies [29, 30]. Then, interviews of the end users, ergo the group members, were carried out to get an overview of the required data. The literature was reviewed to thoroughly understand the data of interest and the relationships across the data objects to ensure data integrity. Relevant literature was found using search words like off-target effects, polypharmacology, and multi-targeting drugs. The collected information was defined in the description of the miniworld, which was to be represented in the database.

2.2.2 Conceptual Design

The next phase in the process was conceptual design. The data requirements were then carefully analysed and described in a conceptual schema. An Entity-Relationship (ER) diagram was used as the graphical notation. A detailed description of the entity types, relationship types, and constraints was defined in the diagram. Each object with independent existence, such as a drug, was represented by an entity. The entity types were then assigned attributes that declared their properties, such as drug name. Each attribute's value domain was defined, imposing restrictions for the data populating process. Finally, one or more key attributes were defined for each entity type. The key attribute was to be populated with a distinct value that could be used to identify each entity uniquely.

Furthermore, the relationships between each entity type were specified. For instance, the relationship type *Interaction* was created between the entity types *Drug* and *Target* to indicate DTIs. Several relationship types were assigned with attributes, for instance, the mechanism of action for a *Interaction* relationship. Additionally, each relationship type was given a cardinality ratio. The constraint defines the maximum number of relationship instances an entity can be part of. For example, a drug can interact with several targets, and a target can interact with several drugs. Hence, the cardinality ratio for the binary *Interaction* relationship type is M:N.

Specialisation was applied to represent binding affinity data. A set of subclasses consisting of the dissociation constant (K_d), the inhibition constant (K_i), and the half maximal inhibitory concentration (IC_{50}) were defined for the entity type *BindingAffinity*, which is the superclass of the specialisation. The specialisation distinguishes how enzyme inhibition is reported. A total constraint was assigned to the specialisation, which means that each entity in the *BindingAffinity* superclass must be member of at least one of the subclasses. Furthermore, the set of entities could be overlapping. To illustrate, one measurement of enzyme inhibition could be reported both as K_i and IC_{50} of the *BindingAffinity* specialisation, in contrast to disjointness where each entity of the specialisation at most can be member of one subclass. In short, the specialisation is overlapping total.

2.2.3 Logical and Physical Database Design

In the next phase, a relational database schema was designed. The high-level data model created from the conceptual design was then translated into a logical design. This process is also known as data model mapping.

Each entity type was mapped to a relation (table) with an associated schema. In the schema, the relation name was defined. The name specifies the entity type, for example, *Drug*. A column was created for each attribute, and the data type was defined. The key attribute was translated to a primary key, representing a unique identifier which could not be NULL. The constraint ensures entity integrity. For example, there is a unique DrugID for each *Drug* entity. In the relation extension, tuples were to be established for each entity, and constraints were defined to ensure that each tuple contained only valid values.

Foreign keys were used to map the relationship across the entity types. If a key did not exist, it was assigned a NULL value. By guaranteeing that all references were valid, reference integrity was achieved. The binary 1:N relationship type representing the relation between a cell line and a disease was mapped by including a foreign key in the *CellLine* relation. Consequently, a cell line could only be associated with a single disease. M:N relations were mapped by establishing relationship relations. These relations consist of the primary keys from the entity types participating in the relationship. The combination of these represent the primary key of the relationship relation. In addition, they work as foreign keys.

A normalisation process was executed to ensure effective data storage in the relational database. This process was essential to prevent redundancy and insertion, deletion, and modification anomalies. Every relation was analysed top-down by reviewing the functional dependencies to evaluate if decomposition was required to avoid storing the same information in several places. The functional dependencies define the semantic constraints across the attributes within a relation schema.

Finally, the database was implemented using the database management system (DBMS), Structured Query Language (SQL). An SQL script was established to construct the database comprised of the relations defined during the logical design, which is provided in the GitHub repository created for this project (https://github.com/KristineLippestad/DrugProfiler). The script specified data types for each attribute in a relation. Furthermore, constraints were included to set the primary keys and foreign keys. To ensure that the foreign key was either updated or deleted if the value was altered in the reference table, ON UPDATE CASCADE and ON DELETE CASCADE were used as foreign key constraints.

2.3 Data Collection

When the database schema was implemented, the database was to be populated with data. Data was collected from multiple acknowledged databases or manually retrieved from the literature to provide comprehensive information on drugs and their targets.

Target Data

Data on potential target genes was collected from the freely accessible resource, Universal Protein Resources KnowledgeBase (UniProtKB). UniProtKB is a comprehensive protein sequence knowledge base that provides data on protein sequences and functional information. The information is integrated from multiple authoritative resources, and the experimental and predicted data is highly curated by experts [51]. The protein data was filtered by taxonomy using taxonomy_id:9606, corresponding to Homo Sapiens. Only reviewed entities were considered, meaning biologists have curated the information provided in each entry. The data was downloaded in a TSV format. The columns were customised to include Entry, Entry Name, Gene Names (primary), and Organism. The first column corresponds to unique UniProt IDs and the second column gives the unique The HUGO Gene Nomenclature Committee (HGNC) symbol. The HUGO committee provides a standardised gene name and symbol for every known human gene [52].

Drug Data

Data on cancer drugs were collected from the Open Target Platform and ChEMBL. The Open Target Platform is a publicly available data knowledge base produced to provide information on drug data identification and prioritisation [53]. Because this project is focused around cancer, a search with the term "Cancer" was conducted on the Open Target Platform to collect data on drugs annotated for this group of diseases. The option "cancer" with the Experimental Factor Ontology (EFO) "MONDO_0004992" was selected. A table representing "Known drugs" is provided on the disease's profile, and this was downloaded as a TSV file. The TSV file includes information about clinical precedence for drugs that are under investigation or approved for cancer treatment. Data on ChEMBL ID, drug name and type was retrieved from this file.

Information about the drugs' chemical structure, indicated by the canonical Simplified Molecular Input Line Entry Specification (SMILES) and The International Chemical Identifier (InChi Key), was collected from ChEMBL. ChEMBL is an open large-scale drug discovery database of small bioactive molecules with drug-like properties [54]. The information was retrieved by executing an advanced search by IDs for drugs collected from the Open Target Platform. The search result was downloaded as a TSV file.

In addition, data harvested from the Open Target Platform was used to establish DTIs with the associated mechanism of action and the action type. Along with the diseases a drug is indicated for, with the reached phase, and how a target and a disease are related.

Drug Sensitivity Data

To provide information on drug sensitivity, data was harvested from Genomics of Drug Sensitivity in Cancer (GDSC). GDSC is a public database providing anticancer drug sensitivity data across cancer cell lines and molecular markers of drug responses. The sensitivity data is generated from high-throughput screening by the Cancer Genome Project at the Wellcome Trust Sanger Institute (WTSI) and the Center of Molecular Therapeutics at Massachusetts General Hospital [55]. From the screening entries, maximal inhibitory concentration (IC_{50}) (μM) and the area under the curve (AUC) were collected. In addition, data on the screened cell line was retrieved, namely: name, tissue type, tissue subtype and the related disease.

Cell Line Ontology (CLO) symbols were identified for every cell line to include unique identifiers. The CLO is created to provide cell line data from numerous resources in a standardised, logically defined format. CLO identifiers were fetched from BioPortal as a CSV file.

Binding Affinity Data

Drug affinity data across human proteins was collected to provide quantitative measurement for DTIs. The data was retrieved from BindingDB, an experimental protein-small molecule interaction database. The database is publicly accessible and holds data collected from scientific articles, US patents, and other databases such as ChEMBL and PubChem. Each entry includes data on one or more protein targets, a drug, associated compound-protein affinity measurements and the data source. In addition, data on key experimental conditions are included for many entries [56].

All data from the resource was downloaded as a TSV file with SMILES strings. Of interest was the quantitative measure of affinity between a drug-target (DT)-pair. Furthermore, experimental conditions, including pH and temperature, when provided. The organism associated with the protein target and curation or data source were also included.

If binding affinity data for an inhibitor of interest was not included in the database, the data was retrieved by searching the literature.

2.4 Application Programs

Application programs were developed to populate the database with data from the used resources. The aim was to create code that enabled the insertion of the collected data and the updation of the database when new data became available. The programs were used to structure and homogenise the data from the various resources before being inserted into the designed framework. Access dates for each resource are shared in Table A.4, in Appendix A.1 - Tools and Data Resources at page 72.

In total, five application programs were developed, and these are listed below:

- An application program to insert data collected from UniProtKB into the *Target* relation.
- An application program to insert data from the Open Target Platform into the *Drug* and *Disease* relations. In addition, it is used to relate *Gene* and *Disease* entries in the *GeneAssociation* relation, to relate *Drug* and *Disease* entries in the *IndicatedFor* relation, and to relate *Target* and *Drug* entries in the *Interaction* relation.
- An application program to update the *Drug* relation with attribute values for Smiles and InChi Key collected from ChEMBL.
- An application program to insert binding affinity data collected from BindingDB and the literature into the *BindingAffinity* relation and associate each measurement with the correct *Drug* and *Target* entries.
- An application program to insert cell line data into the *CellLine* relation, and relate *CellLine* and *Disease* entries. In addition, insert sensitivity data measured across the cell lines into the *Sensitivity* relation.

Figure 2.3 shows the data attributes retrieved from UniProtKB, the Open Target Platform, ChEMBL, GDSC, and BindingDB, and how they are implemented in the database relations. The programs are provided in the GitHub Repository. Python was the applied host language, and sqlite3 the data sub-language. The version number for the utilised programs are shared in Table A.1 and the version of the applied Python packages are given in Table A.2, in Appendix A.1 - Tools and Data Resources on page 72. All developed functions are documented with purpose, their parameters and what they return.

The application programs had to be run in a specific order for the data to be inserted correctly. First, the application program to insert data from UniProtKB was run, and then the application program for the Open Target Platform. They were followed by the application program for GDSC. Finally, the application programs for ChEMBL and the BindingDB were run.

2.4.1 Code Implementation

Functions were created to populate each relation in the database. The general framework for these functions was to first connect to the implemented database file. Secondly, the collected TSV or CSV file with data values was opened for reading before being converted to a pandas data frame. A for-loop fetched data from each entry by iterating over the lines in the data frame. The attribute values were fetched using indexing on the row index and column name. The latter is provided as argument in each function. If needed, the data type was corrected to the format specified in the database design. Some values were also modified to homogenise the data. An Insert operation was executed to create a new tuple in the relation using the identified attribute values. To allow for

future database updates to include new information provided by the data resources, "INSERT OR UPDATE" was used rather than only "INSERT". After executing the insert operation, the change was committed. When the for loop was completed, the database was closed. The framework used in each function developed for insert operations is shown in code example 1.

Listing 1: Example function for insert operations to create new tuples with attribute values for a, b, and c, into relation R using data from a TSV file.

```
import sqlite3
import pandas as pd

def insertTupleIntoRelationR(db_file, data_file, a, b, c):
    """Insert tuples with attribute values for a, b, and c, retrieved
    from a TSV file, into a relation R.
    :param db_file: database db_file, data_file: tsv file from resource,
    a: column name for a in data_file, b: column name for b in data_file,
    c: column name for c in data_file"""

    create_connection(db_file)
    with open(data_file, "r") as data_file:
        df = pd.read_csv(data_file, delimiter="\t")
        for ind in df.index:
            A = df[a][ind]
            B = df[b][ind]
            C = df[c][ind]

            cursor.execute("INSERT OR REPLACE INTO R VALUES (?,?,?)",
            (A, B, C))
            con.commit()

    con.close()
```

In some application programs, additional functionality was required to enable data insertion. In the program developed to insert data from ChEMBL, the attribute values on Smiles and InChi Key were to be inserted into the *Drug* relation, where tuples for each drug already exist. Therefore, an update operation was used to change only these attribute values in each tuple. The data was fetched from the ChEMBL file, and a WHERE clause using the DrugID was used in the query to find the correct drug entry.

Additional functionality was also implemented in the application program developed to fetch data from GDSC. In the *CellLine* relation, CLO was selected as the unique key attribute. Because this data value was not included in the TSV file collected from GDSC, a dictionary was created to identify the CLO for each entry. The dictionary was produced from a CSV file downloaded from BioPortal, representing all CLO and the corresponding synonyms. The ontology ID was used as the key, and the synonyms for the cell line name as values. When inserting data from the GDSC file, the code looked up the correct CLO by searching for the cell line synonym in the dictionary.

Additional drugs were inserted into the database from GDSC. Drug ID and drug type were not defined in the collected data file. To retrieve this information, a dictionary was created, including all the drugs, in addition to separate dictionaries for each drug type. The dictionaries were established from TSV files downloaded from ChEMBL. The ChEMBL IDs were set as keys, and the drug synonyms were set as values. The drug ID for each drug was identified by looking up the drug name from GDSC in the dictionary of all the drugs. Furthermore, the correct drug type was found,

using an if statement to determine which drug type dictionary the drug was included in.

The identifier used in the *Disease* relation was EFO. Therefore, The Cancer Genome Atlas (TCGA) classification used by GDSC had to be translated to EFO to identify the relationship between a cell line entity and a disease entity. The translation was executed utilising a dictionary where the TCGA classifications were used as keys, and their associated EFO were used as values. The TCGA classification from GDSC could then be searched to fetch the correct disease ID.

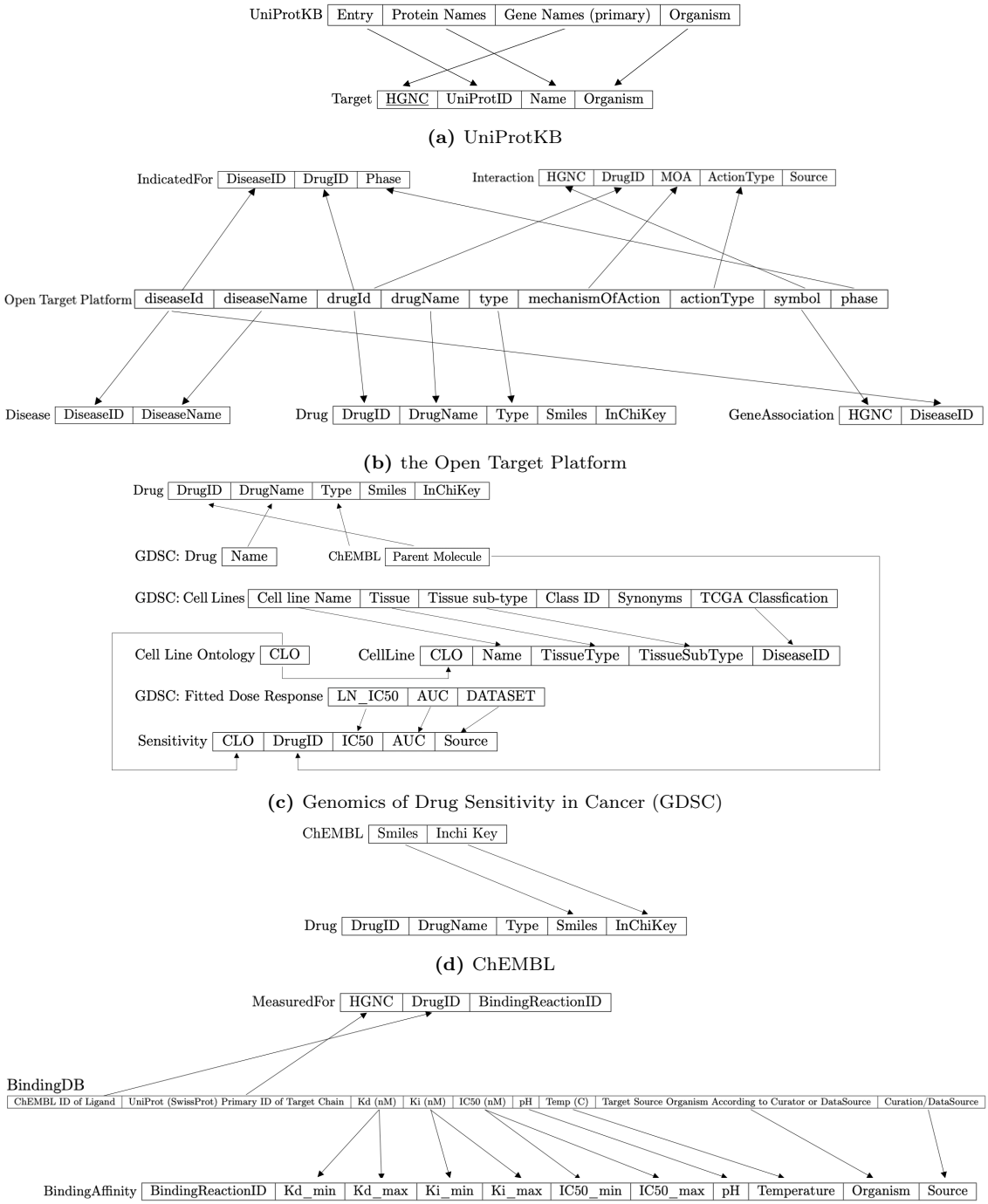
The application program for binding affinity inserts data harvested from a TSV file from BindingDB. In the file, the target in a DTI is given by a UniProt ID. A dictionary was created from the *Target* relation to fetch the corresponding HGNC value for each target, used as the primary and foreign key. In the dictionary, the UniProt ID was implemented as the key and the HGNC symbol as the value. Moreover, some binding affinity values were reported with less than ($<$) and more than ($>$) symbols. The values were translated to numbers to make future queries easier by implementing minimum and a maximum attributes for K_d, K_i, and IC_{50}. An if statement was used to set the min value to be 0 if a less than symbol was used. When a greater than symbol was used, inf was set as the max value attribute. Furthermore, if the binding affinity value was reported as "nan", it was translated to None. When the binding affinity was reported as a regular number this was used both as the min and max value.

Multiple entries in the BindingDB TSV file did not report the drug with ChEMBL ID but with the Smiles or the InChi Key. All three values were therefore used to identify if a drug was included in the *Drug* relation and to find the correct DrugID. If the drug and the target in a binding affinity measurement were stored in the database, a new tuple was inserted in the *MeasuredFor* and the *BindingAffinity* relation.

Unique IDs were created for each entry when inserting binding affinity data into the database because no official identifier for DTIs exists. The IDs were created using the following format HGNC_ChEMBLID_ number. A dictionary was made to provide the correct number in the id when a new measurement was registered for a DTI. If a DTI did not exist in the dictionary, the id got the number 1. Then, a new key-value pair was created for the DTI. The key corresponded to the DTI, and the number corresponded to the number of measurements stored for the pair. When additional measurements for a DTI was inserted in the database, the number of existing records could be fetched from the dictionary. The number was increased by one and added to the ID for the new measurement. Furthermore, the value for the DTI key in the dictionary was increased with one.

A function to insert binding affinity data from the literature was also created. Similar to the functionality to insert binding affinity data from BindingDB, it was confirmed that the drug and the target existed in the database. To create correct IDs for new binding affinity measurements, the functionality verified if a measurement was already registered on the DTI. If no measurements existed, the ID was assigned the number 1. Otherwise, the highest existing number was fetched and increased with one in the created binding affinity ID.

When the database was finalised, it was browsed using DB Browser for SQLite. It was manually verified that the data was inserted correctly by checking multiple random entries in each relation.



(a) UniProtKB

(b) the Open Target Platform

(c) Genomics of Drug Sensitivity in Cancer (GDSC)

(d) ChEMBL

(e) BindingDB

Figure 2.3: Overview of the restructuring process to populate the designed Drug-Target Interaction database with data from a) UniProtKB, b) the Open Target Platform, c) Genomics of Drug Sensitivity in Cancer (GDSC), d) ChEMBL, and e) BindingDB.

2.5 Implementation of DrugProfiler Tool

The DrugProfiler tool was created to make the database content accessible for users that lack extensive programming skills. The GUI allows users to perform queries using menus and forms. The aim was to let the user retrieve customised target profiles for drugs of interest with selected specifications of binding affinity cut-offs and experimental conditions.

The web-based tool is developed using the Dash library based on the Python programming language. The library allows for interactive visualisation of the data of interest. The implemented Python file can be accessed from the GitHub repository. For each function, detailed documentation is provided, stating the purpose, the input parameters, and what is returned. The repository is open to the public, allowing for the download of the file to apply and update the tool.

The tool consists of seven components. A dropdown component was created at the top of the interface to select drugs of interest. The given alternatives were restricted to drugs with associated binding affinity measurements. On the left-hand side of the interface components to select the conditions for the queries were implemented. Including sliders to specify the binding affinity cut-offs for K_d, K_i, and IC_{50}. Below, input containers for experimental conditions for the binding affinity measurements, including temperature and pH, were created.

On the right-hand side of the webpage, containers to present the query results were made. A text box was created in the first container to show the generated target profiles. The profiles are produced with queries, defined to fetch targets where binding affinity was measured below the set cut-offs. Furthermore, if pH and temperature were specified, the records were filtered according to these conditions. Functionality was created to translate the query result to text, represented in this container.

In addition, code was developed to make a .txt file with a format corresponding to the drug panel input file used in the DrugLogics Pipeline. The .txt file is downloaded when a button named "download-drug-panel" is pushed. Next to the button, an input component is provided, where the downloaded file name can be specified.

A network figure illustrating the generated target profiles was implemented in the container below using the Dash Cytoscape package. Radio items were created to enable the selection of the drug for visualisation. Nodes were established for the selected drug and its targets, and edges were made to represent the interaction between them. A query was defined to determine the interval between the stored binding affinity values for each DTI, below the set cut-offs. Node size was defined according to the lowest binding affinity value stored for the interaction. To decrease the probability for False Positive (FP) and False Negative (FN) DTIs, it is recommended to integrate bioactivity data from numerous studies [57]. For this reason, code was created to colour the target nodes light or dark grey to indicate if one or multiple binding affinity values were measured for the DT-pair, respectively.

Furthermore, functionality was defined so the user could press a node for additional information about the DT-pair. The information includes the interval for the measured binding affinity values and the standard error of the mean. The latter is included to estimate the variability across the binding affinity measurements for the selected DT-pair.

Functionality was implemented to allow users to compare target profiles identified for various drugs. A checklist of the selected drugs was made. A function was then defined to identify mutual targets between drugs checked off to be compared. The output result is converted to text. Furthermore, it is used as input in a function that creates a network figure where nodes for the shared targets and the drugs are generated, along with connecting edges.

Finally, a container providing sensitivity data was created. Radio items were implemented for the selected drugs if sensitivity data existed. Furthermore, a drop-down with cell lines was made. The alternatives were filtered only to include cell lines with sensitivity data for the selected drug. A query was then performed to fetch the drug sensitivity data across the selected cell lines. The result is translated into descriptive text.

2.6 Drug Target Profiling

In this project, a panel of 19 drugs were chosen for analysis, which earlier have been studied by the DrugLogics initiative [29, 58, 30]. The selected drugs inhibit signalling proteins that participate in established signalling pathways, such as PI3K/AKT, NFkB, JAK/STAT, CTNNB1/TCF, and MAPK pathways. Two drugs, SF1670 and GSK 429286, were excluded from the analysis as adequate binding affinity data could not be identified the applied database or in the literature. The key information for each drug is provided in Table A.5, in Appendix A.2 - Chemical Inhibitores on page 74.

Target profiles for the remaining 17 small-molecule inhibitors were successfully identified and characterised using the DrugProfiler tool. The profiles were generated using four different cut-offs, to enable evaluation of the interaction strength for the identified DT-pairs. The applied cut-offs for K_d, K_i, and IC_{50} were 10 nM, 100 nM, 1 µM, and 10 µM. The generated profiles are accounting for strong and moderate interactions. The highest cut-off was set to < 10 µM, as affinity above this cut-off is considered weak [59]. The query result was downloaded as text files.

The generated target profiles made the foundation for the further study of how off-target effects alter predicted drug response. Prior to conducting the modelling, a thorough analysis of the target profiles was performed to evaluate selectivity.

2.6.1 Literature of Mechanism of Inhibition

Drugs that showed binding affinity toward numerous off-targets were evaluated through a literature study. The objective was to investigate the potential impact the mechanism of action had on selectivity. For each drug, it was identified how the DTI occurred, including if the inhibitor competed with the enzyme's substrate and the specific region of the enzyme the inhibitor bounds.

2.6.2 Analysis of Target Composition with KinMap

The next objective was to investigate the distribution of drug binding evidence across protein targets of the human kinome. The study was accomplished by creation of a phylogenetic trees of the human kinome for each drug, annotated with the identified targets. The kinome tree was provided by the web-based tool KinMap, which allows for an interactive study of the human kinome [60].

The annotation was conducted according to the generated target profiles. Only kinase inhibitors and kinase targets were considered. Each annotation was labelled with the kinase name and a circle. The size of the circle was scaled based on the binding affinity data. Hence, four different dimensions were used according to the applied cut-offs. If a target name was ambiguous, the correct name was specified among the kinase name suggestions or excluded if not present in the map.

2.7 Drug Response Analyses to Evaluate the Impact of Off-Target Effects

The overarching objective of this project was to use logical modelling to evaluate how off-target effects influence the prediction of drug response. Stochastic modelling was conducted to simulate how cell growth was inhibited when off-target effects were included compared to only accounting for

the primary targets. Subsequently, the DrugLogics pipeline was utilised to predict drug synergies with and without including off-target effects in the perturbation data, aiming to assess the potential improvement in prediction performance.

2.7.1 The CASCADE Models

Two logical models were used in the drug response analyses - namely, CAncer Signaling CAusality DatabasE (CASCADE) 1.0 and 2.0. The prior knowledge networks (PKN) are developed by the DrugLogics Initiative and illustrate cancer signalling events. The first topology, CASCADE 1.0, describes the cell fate decision network in the adenocarcinoma cancer cell line (AGS) gastric cancer cell line. The logical model is developed from a complex signalling and regulatory network built from general signal transduction knowledge. The network is extended with nodes targeted by small molecule inhibitors, studied in drug synergy predictions, and components of general oncogenic processes. The final model includes 75 signalling and regulatory network nodes connected by 149 directed interactions [29]. CASCADE 1.0 is presented in Figure 2.4a.

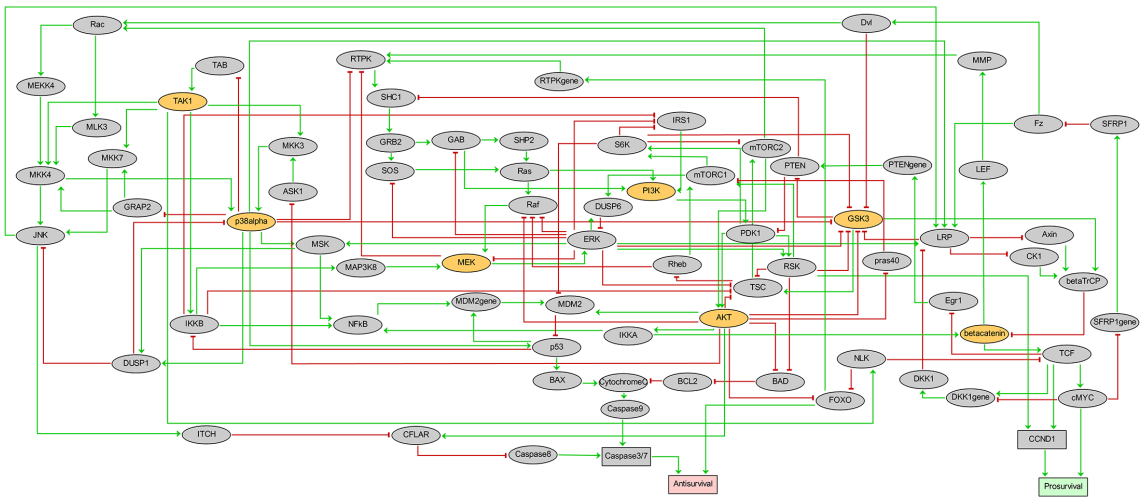
CASCADE 2.0 is an extension of the first model representing a more generic cancer model which includes new network components relevant to generic signalling pathways and target nodes of additional inhibitors. The extension was carried out by manual curation using the knowledge databases Signor [61], KEGG [62], and by recent scientific literature. CASCADE 2.0 comprises 144 nodes and 367 interactions [30]. CASCADE 2.0 is presented in Figure 2.4b.

In both models, the activity state of each node is expressed by a Boolean value, 0 if inactive and 1 if active. A selection of nodes are multi-valued, which allows for the modelling of graded growth or inhibitory effects. Among the multi-level nodes are the output nodes that state cell fate phenotype. These are named *Prosurvival* and *Antisurvival*, and each can possess four levels (0, 1, 2, 3). In addition to the immediate upstream nodes - namely, Caspase 3/7 and CCDN1 (0, 1, 2). Logical rules are defined for each network node translated from baseline activity data from a proliferative state. The associated rule determines how the activity level is altered by the rules defined for its regulator [29]. The logical models are provided as GINsim files in the GitHub repository.

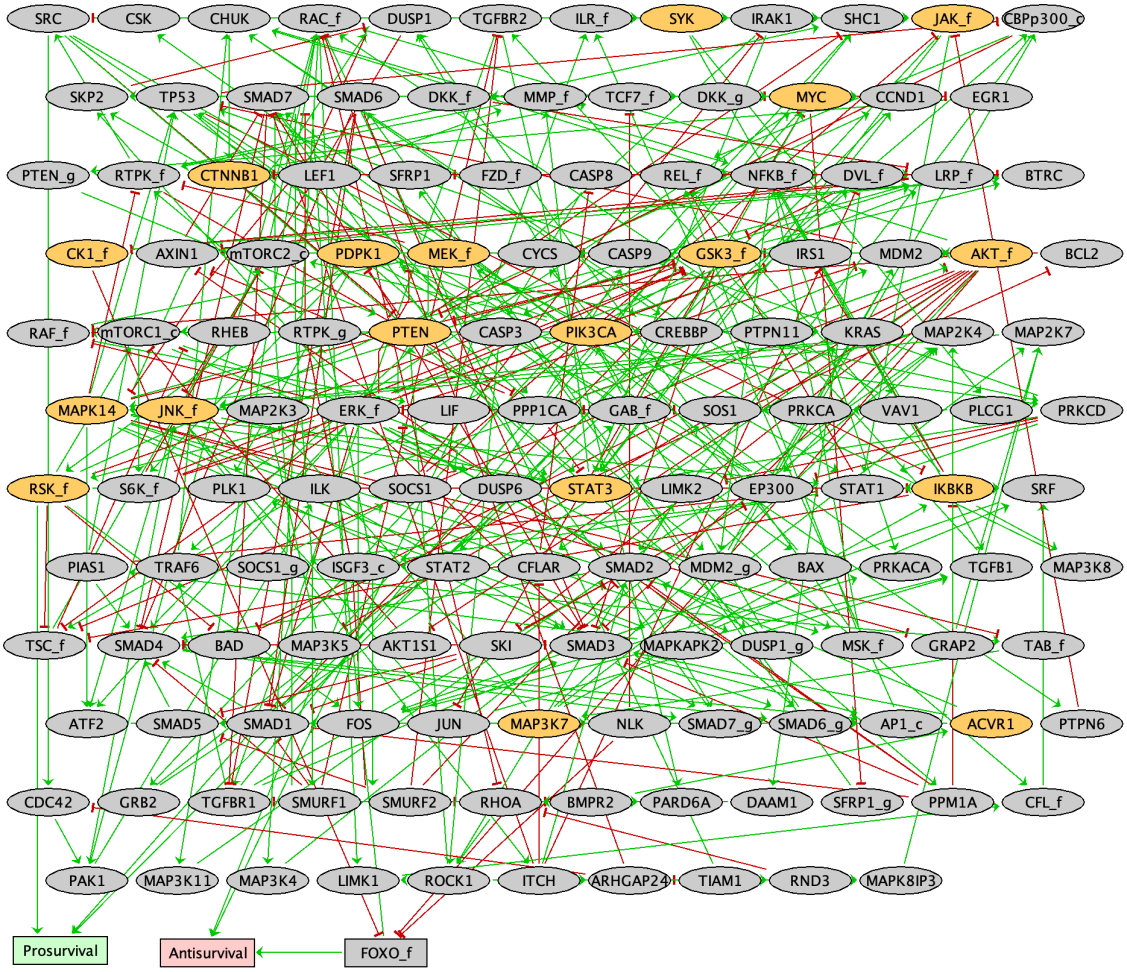
2.7.2 Filtering of Target Profiles

Before performing drug response simulations, the generated target profiles were adapted to only include targets present in the CASCADE models. In addition, the target names were translated from the HGNC symbol to the node name used in the logical models.

A Python script was developed to perform the filtration and translation automatically, which is provided in the GitHub Repository. The drug panel text files downloaded from the DrugProfiler tool were updated utilising the script. The file content was adapted according to dictionaries created for each of the logical models. The node names were used as keys, and the corresponding HGNC symbols were used as values. The script iterated through each line in the text file and edited the target name. The correct name was fetched by looking up the value (HGNC symbol) and then getting the associated key (node name). Targets that are not included in the model were excluded. The function then called a helper function that writes the updated target profiles to a new text file.



(a) CASCADE 1.0



(b) CASCADE 2.0

Figure 2.4: Directed graphs showing the logical models a) CAncer Signaling CAusality DatabasE (CASCADE) 1.0 and b) CASCADE 2.0. The model includes two output nodes, namely *Antisurvival*, coloured red, and *Prosurvival*, coloured green. The red blunted arrows indicate inhibiting edges and the green pointed arrows indicate activating regulations. Each network component is assigned by a Boolean variable (0 or 1) or a multilevel variable, denoted by a ellipse or rectangle, respectively. Primary targets of the analysed drugs are marked yellow.

2.7.3 Prediction of Cell Fate Phenotype

Alterations in cell viability were predicted using the Markovian Boolean Stochastic Simulator (MaBoSS) to investigate how off-target effects influence drug response. MaBoSS is a software environment that performs stochastic simulations on a Boolean network, which enables the modelling of signalling pathways behaviour in response to a drug treatment [63]. The simulations were carried out in the CoLoMoTo interactive notebook. An environment unifies multiple software tools to edit, conduct, share, and reproduce analyses of qualitative models of biological networks, including MaBoSS [64].

In the stochastic simulations, the Boolean networks are first converted into continuous time Markow processes. The activation and inactivation of every network component are linked to an up and a down rate, which indicates the propensity of the corresponding transition. Starting from the defined initial states, the simulation evaluates all potential node updates, and calculates the likelihood and duration of each change. Time trajectories and the probability of a model state for the defined set of initial conditions are generated for the entire simulation time. Consequently, the distribution of steady states and the probability for each can be estimated [65].

The Boolean networks, CASCADE 1.0 and CASCADE 2.0, were simulated. The models were downloaded as GINsim files from the GINsim model repository and from the supplementary material by Niederdorfer et al. named "AGS_literature-informed.zginml", respectively [29, 30]. Before running the simulations, the models were exported from a GINsim format to a MaBoSS format. The initial states were defined by setting the three prosurvival nodes representing the levels of prosurvival to be active (state = 1) to reflect proliferating cancer cells. To predict the cell fate phenotype, the node state of the six output nodes representing the levels of prosurvival and antisurvival had to be determined. Therefore, these nodes were selected as the output nodes, while the remaining nodes were ignored.

The first simulation was conducted without any perturbations to determine the cell fate phenotype of the wild-type model. Subsequently, simulations of the response to seven small molecule inhibitors studied by Flobak et al. (2015) were conducted using CASCADE 1.0 [29]. Followed by simulations with CASCADE 2.0 using 17 small molecule inhibitors studied by Niederdorfer et al. [30]. Perturbations were introduced according to the filtered target profiles. The inhibition of each target was modelled by fixing the activity of the target nodes to 0. Only the primary targets were inhibited in the first simulation to enable the identification of alterations when off-target effects were considered. Subsequently, the model was perturbed using the four target profiles generated for each drug using different cut-offs.

After each simulation, the output was visualised with a pie chart representing the asymptotic distribution of probabilities of reaching each stable state at the last time point. The predicted stable states comprise the output nodes representing the prosurvival and antisurvival levels. These were applied to calculate a growth sum on a scale from -3 to +3 using the formula 1, to define the cell fate phenotype.

$$Cell\ growth = \#\ active\ prosurvial\ nodes - \#\ active\ antisurvival\ nodes \tag{1}$$

The implemented notebook is provided in the GitHub repository.

2.7.4 Prediction of Drug Synergies with the DrugLogics Modelling Pipeline

To evaluate if consideration of off-target effects could improve the prediction performance of synergistic drug combinations, simulations using the DrugLogics software pipeline were executed.

The software module was installed using the Docker image "bblodfon/druglogics-synergy:1.0" (*1.0* version, *druglogics-synergy* at *1.2.1* version), which includes all dependencies and external libraries. The druglogics-synergy repository was then cloned from git, fetching the input directories.

The simulations were executed using the ags_cascade_1.0 and the ags_cascade_2.0 directory. Each directory consists of multiple input files. The first file is a single-interaction network file, *network.sif*, that describes the CASCADE 1.0 and CASCADE 2.0 topologies, respectively.

Secondly, *training* files were used to calibrate the general models. All simulations were first conducted using a file with baseline activity data collected from the literature by Flobak et al. 2015, where the model was trained to become AGS specific [29]. Thereafter, the simulations were repeated using a file with data matching a proliferation profile where the system is left unperturbed, which correspond to a direct reflection of a cancer cell network system, to determine if the predictions could be improved by normalisation.

A *modeloutputs* file, is applied to define the network nodes that signal cell proliferation and apoptosis when active. These nodes are used to identify the model's attractors to compute the output growth response. For simulations with CASCADE 1.0, cell proliferation was signalled when the three nodes RSK_f, MYC and TCF7_f were active, and apoptosis was signaled by CASP8, CASP9 and FOXO_f. The file was identical for CASCADE 2.0, except from TCF_f that was replaced by CCND1.

For each of the network models, simulations were carried using five distinct *drugpanel* files. The input file provides the drugs to be analysed and their associated target profile. The first simulation was carried out using the primary targets exclusively. Then adapted drug panels generated for the cut-offs 10 nM, 100 nM, 1 µM and 10 µM were used as input. Despite the lack of identified targets below a set cut-off, the primary targets were included in all target profiles. The primary targets were included to cover all the drugs in every simulation.

The final input file, *config*, states the parameter applied by software modules. Highest Single Agent model (HSA) was selected as the model used to calculate drug synergies. According to the model, a drug combination is classified as synergistic if a combination's effect is higher than that observed from any of the single drugs alone [66]. The HSA was calculated according to the formula 2.

$$Viability(Drug\ A\ + Drug\ B)\ -\ min[Viability(Drug\ A,\ Drug\ B)] \tag{2}$$

Complete documentation of each file is provided in the DrugLogics Software Documentation.

For each simulation, a new directory was generated inside the respective folder where the produced output files were saved. The output files include the Boolean models produced by the genetic algorithm and the synergy predictions. The output results were analysed using the emba library in R [67]. The R script used to analyse the synergy predictions was fetched from "Tutorial for synergy prediction using the DrugLogics software pipeline". The script was adapted to analyse and download the synergy predictions and plots indicating the prediction performance generated from various drug panel, simultaneously. The updated script is provided in the GitHub repository.

The predicted synergies were structured in a table with the calculated synergy score and an associated 1 or 0, indicating whether the synergy was observed experimentally in drug screens performed by the DrugLogics initiative [29, 58]. The model performance was evaluated to determine whether accounting for off-target effects improves or weakens the predictions. The output predictions were characterised as True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) based on the observed synergies to analyse the performance. Next, the characterisation results were used to compute the True Positive Rate (TPR), also known as sensitivity or recall,

to determine the probability of an observed synergy being predicted by the software, using equation 3. Then, the False Positive Rate (FPR) was computed to determine the probability of a non-synergistic combination to be predicted as positive by equation 4. Finally, the precision was calculated to determine how many predicted synergies were observed by equation 5.

$$True\ Positive\ Rate = TP/(TP\ +\ FN) \tag{3}$$

$$False\ Positive\ Rate = FP/(TN\ +\ FP) \tag{4}$$

$$Precision = TP/(TP\ +\ FP) \tag{5}$$

The calculated values were used to plot receiver operating characteristic (ROC) curves, by plotting the True Positive Rate (TPR) against False Positive Rate (FPR). In addition, precision-recall (PR) curves were generated by plotting the precision against the recall. The AUC value was calculated for each plot, which was used to rate the prediction performance.

Finally, the synergies scores obtained from the steady state calibration were normalised by subtracting the random model predictions. The normalised prediction scores were used to calculate updated ROC and PR AUCs.

2.8 Use of the DrugProfiler Tool in the Selection Process of Cancer Treatment

The final challenge taken up in this project was to perform analysis with the DrugProfiler tool in the process of selecting an effective drug for a patient enrolled in the IMPRESS-Norway study. This is a precision medicine study focusing on cancer. Patients with a complex cancer diseases that no longer benefit from standard treatment can take part in the study. A gene panel analysis of the patient's tumour is performed to characterise 523 genes. The obtained molecular profile is used to select the cancer treatment. The study offers off-label treatment, using drugs that are approved for particular cancer diseases on new types [68].

The DrugProfiler tool was used to evaluate if a patient could benefit from treatment with off-label use of *Imatinib*. A target profile was generated for the drug, and the overlap between the identified targets and the mutations found in the molecular profile was analysed. Furthermore, the binding affinity strength towards the targets of interest was assessed.

3 Results and Discussion

3.1 The Drug-Target Interaction Database

The overarching objective of this project was to be able to study how off-target effects of so-called targeted drugs may affect model-based predictions of drug responses. For this, a database application was developed to enable efficient handling of such off-target effects. The application was named the Drug-Target Interaction database, and it consists of a relational database and associated programs developed to execute database updates and queries. The database contains cancer drugs, targets, and the prior knowledge available on drug-target interactions. In addition, supplemental information is stored to allow for a thorough investigation of a drug's target profile.

3.1.1 Defined Data Requirements

The database design process was initiated by a requirement collection and analysis to pinpoint the data requirements of the prospective end users of the Drug-Target Interaction database. The main focus was on the DrugLogics participants, in addition to systems biologists, bioinformaticians and other scientists working with polypharmacology. After completing this step, the following miniworld, describing the components provided in the database, was defined.

- A drug-target interaction (DTI) comprises a drug and a target. Each drug has a unique identifier retrieved from ChEMBL, a name, a type, and a molecular structure described by Simplified Molecular Input Line Entry Specification (SMILES) and The International Chemical Identifier (InChi Key).
- A drug interacts with one or more targets, each described by a unique HGNC gene symbol from the HUGO Gene Nomenclature Committee, a unique identifier from UniProtKB, a complete name, and the source organism.
- The database keeps track of a DTI's mechanism of action, the action type and the source of the interaction.
- A drug's target space may be analysed quantitatively by binding affinity measurements. Every measurement has a unique identifier. The binding affinity is expressed by the dissociation constant (K_d), inhibition constant (K_i), or half maximal inhibitory concentration (IC_{50}). If accessible for a measurement, the experimental details for pH and temperature are provided.
- A drug can be indicated for one or more diseases, and the database stores each disease's Experimental Factor Ontology (EFO) identifier and the name. Furthermore, each indication is provided with the phase reached in clinical trials. A disease can also be associated with multiple targets.
- Sensitivity data may be stored for a drug, screened across cell lines. A cell line can be derived from a patient with a specific disease. Each cell lines have a unique Cell Line Ontology (CLO), a name, tissue type, and tissue subtype. The sensitivity is reported by the half maximal inhibitory concentration IC_{50} and the area under the curve (AUC) for each experiment.

The first entity type that keeps track of DTIs consists of qualitative data curated from the scientific literature and post-marketing package inserts. The data allow for straightforward interpretation, considering only confirmed targets. However, quantitative data was required to enable a more

extensive exploration of a drug's target space and to prioritise the components. Binding affinity measurements were therefore included as a new entity type, capturing potential interactions across numerous proteins.

Although binding affinity is of high value when examining the target space, some sources of uncertainty in affinity measurements should be considered when operating with this data. When a binding affinity value is defined experimentally, a "signal" is measured as a function of the concentration of the ligand (drug) and/or protein. This signal can, for example, be a temperature change when isothermal titration calorimetry is applied or altered substrate or product concentration when using enzyme inhibition assays. Wrong or inconsistent signal measurements can arise from instrumentation issues, leading to the incorrect concentration of free ligands. Furthermore, inaccuracy in the actual or effective concentration of free protein or ligand (drug) can appear. Errors can appear if the ligand has a chiral center because the binding constants can differ considerably between the enantiomers. A protein can also have more than one binding site for the ligand, and using one-to-one binding math will result in incorrect calculations [69].

Nevertheless, data provided in the database is detected under various experimental settings, enzyme substrate when using an enzyme IC_{50} assays, and cosolutes [69]. All these factors can cause variation in the data. Despite these possible sources of uncertainty, the probability for False Positive (FP) and False Negative (FN) DTIs is decreased when bioactivity data is integrated from multiple studies [57].

Next, a relationship between the entities for drug and disease was created, with an associated attribute stating the phase reached in clinical trials. This information might not be relevant if an agent is used in research, for instance, to analyse protein function. On the other hand, when the data is applied to select cancer treatment, this is a valuable attribute to identify possible drug candidates.

Finally, in vitro drug sensitivity data were included. The sensitivity data show a drug candidate's efficiency in inhibiting a cell line's growth. This quantitative data facilitates the prioritisation and selection of efficient drug candidates and the required drug dosage when performing analysis on a specific cell line. Even more valuable regarding the context of target profiling, the quantitative data can be used to select the cut-off for binding affinity data to retrieve targets interacting with the drug at the applied drug concentration.

3.1.2 Conceptual Database Design

To accurately describe the defined data requirements a conceptual schema was designed based on the miniworld. The Entity-Relationship (ER) model was applied for conceptual schema design. The ER model is shown in Figure 3.1.

Basic modelling concepts were mainly sufficient to illustrate the database schema for the Drug-Target Interaction database. Entity types, indicated by rectangles, describe the five objects stored in the database. Each entity type has associated attributes with describing properties and a unique key attribute indicated by an underline. The relationship across the entity types is described by relationship types, indicated by a diamond with lines to the participants. A cardinality ratio is defined for each relationship type, defining the maximum amount of related entities. Most relationship types are binary relations where the cardinality ratio is M:N. Hence, an entity can participate in zero or more relationship instances. The exceptions are the *DerivedFrom* relationship, where a *CellLine* maximum can be related to one disease. In addition, the *MeasuredFor* relationship, where a binding affinity measurement must be related to both a drug and a target entity. Furthermore, the *MeasuredFor* relationship type relates three entity types and is therefore ternary.

Finally, a specialisation was used to distinguish the various methods of expressing binding affinity indicated by a circle. Three subclasses were created as specialisations of the superclass *BindingAffinity*. The set of subclasses was $\{K_d, K_i, \text{and } IC_{50}\}$. An entity must be a member of at least one subclass but may be a member of multiple. Consequently, the specialisation is total and overlapping, indicated by a double line to the superclass and an **o** implemented in the circle, respectively.

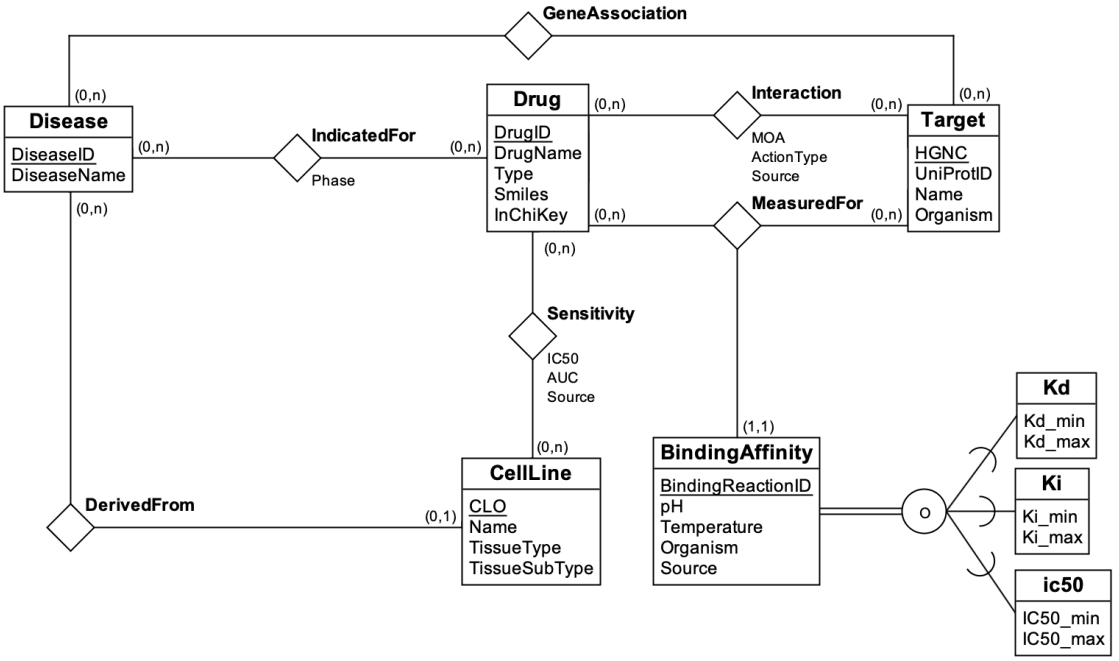


Figure 3.1: Entity-Relationship (ER) schema diagram of the Drug-Target Interaction database. Rectangles represent entity types, each having an entity name and attributes. The key attribute is marked with an underline. Relationship types are presented with diamonds with lines to the participating entity types. The specialisation is represented by a circle with an **o**, indicating that an entity may be a member of more than one subclass. A double line is drawn to the superclass, symbolising the total constraint, which indicates that a *BindingAffinity* entity must be a member of at least one subclass. A line is drawn to each subclass, with a subset symbol stating the direction of the specialisation.

3.1.3 Logical Database Design

The following step in the database design was the actual implementation of the database. This step starts with logical design and results in a database schema comprised of a set of relations mapped from the ER schema (Figure 3.1).

The five entity types are mapped to relations in the database schema. Each relation includes a primary key, indicated by an underline, and the associated attributes. The specialisation for *BindingAffinity* is mapped to one superclass, including the set of subclasses.

The *DevelopedIn* relationship type, with a 1:N cardinality type, is represented by the foreign key DiseaseID in the *CellLine* relation that references the *Disease* relation. The remaining relationship types had N:M cardinality types. These are mapped to relationship relations that cross-reference the primary keys from the related entity types. In these relationship relations, the two referencing attributes are primary and foreign keys.

The resulting database schema is provided below. The primary key, the referencing and the unit of measurement is specified below each relation.

The designed database schema was used to implement the database applying SQL. The SQL script is provided in the GitHub repository. In the script, each attribute is associated with a value set and constraints defining if a value has to be unique and if it can be NULL. Furthermore, constraints specify the primary keys in each relation and the relation a foreign key reference. The Target relation is named Gene in the provided script and database file. However, it is renamed in the report to describe the relation content better.

Drug (DrugID, DrugName, Type, Smiles, InChiKey)

- DrugID is the primary key of the *Drug* relation.

Target (HGNC, UniProtID, Name, Organism)

- HGNC is the primary key of the *Target* relation.

Disease (DiseaseID, DiseaseName)

- DiseaseID is the primary key of *Disease*.

CellLine (CLO, Name, TissueType, TissueSubType, DiseaseID)

- CLO is the primary key of the *CellLine* relation.
- DiseaseID is a foreign key, referencing the *Disease* relation.

BindingAffinity (BindingReactionID, Kd_min, Kd_max, Ki_min, Ki_max, IC50_min, IC50_max, pH, temperature, Organism, Source)

- BindingReactionID is the primary key of the *BindingAffinity* relation.
- The unit of measurement for Kd_min, Kd_max, Ki_min, Ki_max, IC50_min, IC50_max is nanomolar.
- The unit of measurement for temperature is celsius.

Interaction (HGNC, DrugID, MOA, ActionType, Source)

- The combination of HGNC and DrugID is the primary key of *Interaction*.
- HGNC is a foreign key, referencing the *Target* relation.
- DrugID is a foreign key, referencing the *Drug* relation.

MeasuredFor (HGNC, DrugID, BindingReactionID)

- The combination of HGNC, DrugID, and BindingReactionID is the primary key of *MeasuredFor*.
- HGNC is a foreign key, referencing *Target*.
- DrugID is a foreign key, referencing the *Drug* relation.
- BindingReactionID is a foreign key, referencing the *BindingAffinity* relation.

GeneAssociation (HGNC, DiseaseID)

- The combination of HGNC and DiseaseID is the primary key of *GeneAssociation*.
- HGNC is a foreign key, referencing the *Target* relation.
- DiseaseID is a foreign key, referencing the *Disease* relation.

Sensitivity (CLO, DrugID, IC50, AUC, Source)

- The combination of CLO and DrugID is the primary key of *Sensitivity*.
- CLO is a foreign key, referencing the *CellLine* relation.
- DrugID is a foreign key, referencing the *Drug* relation.
- IC50 is the natural log of the fitted IC50, and the unit of measurement is micromolar.

The database schema design ensures minimum redundancy and information preservation, resulting in efficient data storage. The information is divided into smaller tables with fewer attributes and higher normal form. As a result, all information is only stored once, which reduces data redundancy. The alternative is larger tables with numerous attributes which store the same data multiple times. An illustration is a drug-target interaction relation where every drug and target attribute is provided each time they are included in an interaction tuple. This storage strategy would lead to less complicated queries because fewer JOIN operations would be required to retrieve the data of interest. However, the same information would be stored in multiple places resulting in redundancy. Consequently, database inconsistency would quickly arise. Database inconsistency is when the same data appear in various formats in multiple tables [50]. For instance, if the InChi Key value of a Drug was to be changed, and the drug was included in 20 DTIs, a mistake could easily appear when the same value had to be updated in 20 different tuples. In comparison, the value would only need to be updated in one place with the applied database design where the DTI data is split between relations. Under those circumstances, data consistency is easier to maintain and significant storage space is saved.

In addition, the applied design prevents undesired update anomalies. Update anomalies are defined as inconsistency in the data caused by insertion, modification and deletion anomalies [50]. To demonstrate, the alternative DTI storage strategy presented in the preceding paragraph can not accommodate the storage of a target that does not interact with a registered drug. One solution is to use NULL as the attribute values for the drug. However, this approach would violate the entity integrity because the primary key, DrugID, cannot be NULL. Hence, insert anomalies arise. Secondly, modification anomalies arise from the requirement of updating one attribute value in multiple tuples, as illustrated in the InChi Key example. Finally, deletion anomalies could occur. To exemplify, if a target is included in one interaction only and this tuple is deleted, information about this target would be lost.

To conclude, the applied design, with few attributes and lower normal form, mitigates anomaly and redundancy problems, ensuring effective data storage.

3.1.4 Database Content

When the design process was finalised, the database was populated with DTI data collected from multiple resources. Data from each resource was inserted into the database using the five application programs developed for data insert and update. The implemented database file is named

DrugTargetInteractionDB.db and can be downloaded from the GitHub repository. To examine the content in the database file the data was browsed using DB Browser for SQLite. The content can also be viewed using online tools, such as the SQLite Viewer.

Because binding affinity for one measurement most often is expressed by either K_d, K_i, or IC_{50}, and not all three, mapping the specialisation of *BindingAffinity* to one superclass relation resulted in multiple NULL values for the specific attributes in the subclasses. However, managing the NULL values in the queries was more straightforward than using numerous JOIN operations to fetch the data of interest. Furthermore, it was observed that multiple tuples in this relation missed attribute values for pH and temperature. These values were not provided in the data file from BindingDB. Missing information about the experimental conditions contributes to more uncertainty when comparing measures performed for the same DT-pair.

The relations with the highest number of tuples store experimental data, namely *BindingAffinity* and *Sensitivity*. Multiple binding affinity measurements are often given for one DT-pair. Furthermore, in the *Sensitivity* relation, a drug is screened for numerous cell lines. An overview of the number of tuples in each relation is given in Table 3.1.

Currently, the drugs included in the database are annotated for cancer, the disease under study in this project. However, the designed framework supports including drugs annotated for other diseases. Insertion of new records can easily be carried out with the developed application programs for inserting and updating data, following the procedure described for cancer drugs in Materials and Methods on page 18. Currently, the available data allows for the repurposing of cancer drugs to treat other diseases. However, it is with this version of the Drug-Target Interaction database not yet possible to identify drugs annotated for other diseases than cancer for suggesting its use in cancer. Incorporating additional drugs would enhance the database's applicability for studying a broader range of illnesses. Supplementary agents would provide new efficacious candidates for non-therapeutic research settings where the therapeutic objective of a drug is negligible. Additionally, it would enable the exploration of repurposing drugs annotated for diseases other than cancer.

To maintain the content up to date, new data should be retrieved and loaded into the database when new information is provided by the applied data resources. Updating the database is an easy process with the help of the application programs. The only requirement is to download the data files with new information and provide these as input in the implemented functions.

Moreover, additional data resources can be explored to investigate the possibilities for extending the database content. For instance, identifying resources that offer binding affinity data for additional DTIs could prove beneficial in covering new potential targets. Covering more of a drug's target space could enable enhanced predictions of drug effects.

Finally, the database's range of applications can be expanded by introducing new entity types, relationship types, and attributes. For instance, it is possible to include distinct approaches to scale DTIs. An option is to include data from gene expression profiles where measurements are performed before and after drug treatment to provide information on transcriptional responses of drugs. Such data could be retrieved from The Library of Integrated Network-Based Cellular Signatures (LINCS). The LINCS program aims to establish a network-based understanding of human biology by recording alterations in, for instance, gene and protein expression when cells are exposed to a wide range of perturbing agents [70].

To summarise, the final database comprises extensive information that allow for a thorough study of drug, targets, and the interactions between DT-pairs. It is possible to extend the content with new tuples in the existing entity and relationship entries, as well as expand the miniworld represented in the database.

Table 3.1: The number of tuples in each relation in the Drug-Target Interaction database.

Relation	# of tuples
Drug	1567
Target	20159
Disease	405
CellLine	836
BindingAffinity	54491
Interaction	4152
MeasuredFor	54491
GeneAssociation	20790
Sensitivity	87431

3.2 The DrugProfiler Tool

The following task was to develop the DrugProfiler tool linked to the Drug-Target Interaction database for generating and examining drug target profiles. The tool facilitates exploration of the database content, and utilisation does not require a substantial computational background. Multiple inputs can be specified to fetch DTI data of interest. Furthermore, the tool includes a Graphical User Interface (GUI) to represent the retrieved data with text and figures, allowing for a comprehensive interpretation of the result. The overview of the main features of the DrugProfiler tool is shown in Figure 3.2.

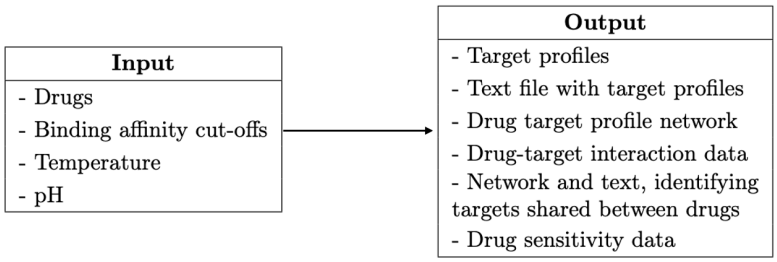


Figure 3.2: Summery of key features of the DrugProfiler tool.

3.2.1 Data Input

The DrugProfiler tool allows users to perform customised queries by adapting the various input values. Figure 3.3 shows the input alternatives provided in the interface. A drop-down box is implemented to select one or more drugs by searching for the drug name or ChEMBL ID (Figure 3.3a), which enables bulk queries for efficient data retrieval. The drug targets are generated according to binding affinity. If binding affinity for a DTI is measured below a set cut-off, the molecule will be defined as a target. The affinity limit values are defined using sliders presented in the binding affinity cut-off container. There are, in total, three sliders, representing K_d, K_i, and IC_{50}, allowing the user to determine the method for binding affinity measurements. The text below each slider states the selected cut-off (Figure 3.3b). The experimental conditions container includes boxes to specify pH and temperature (Figure 3.3c) for the affinity measurements. Finally, the Drug Sensitivity Container includes radio items and a drop-down menu to select a drug and the cell lines for which the user would like to retrieve sensitivity data (Figure 3.3d).

The application offers the flexibility to customise experimental conditions and define binding affin-

ity values to meet the specific query objectives. For instance, if the aim is to achieve the highest possible agreement across measurements, the query can be limited to only include measurements where the affinity values are computed by IC_{50}. The pH and temperature can also be specified to ensure consistent experimental conditions. Another approach is to include all expression options for binding affinity but adjust the IC_{50} cut-off to be slightly lower since IC_{50} values are generally higher than K_d and K_i. If the objective is to decrease the probability of FP and FN, all cut-offs can be set to the same value to include as many measurements as possible.

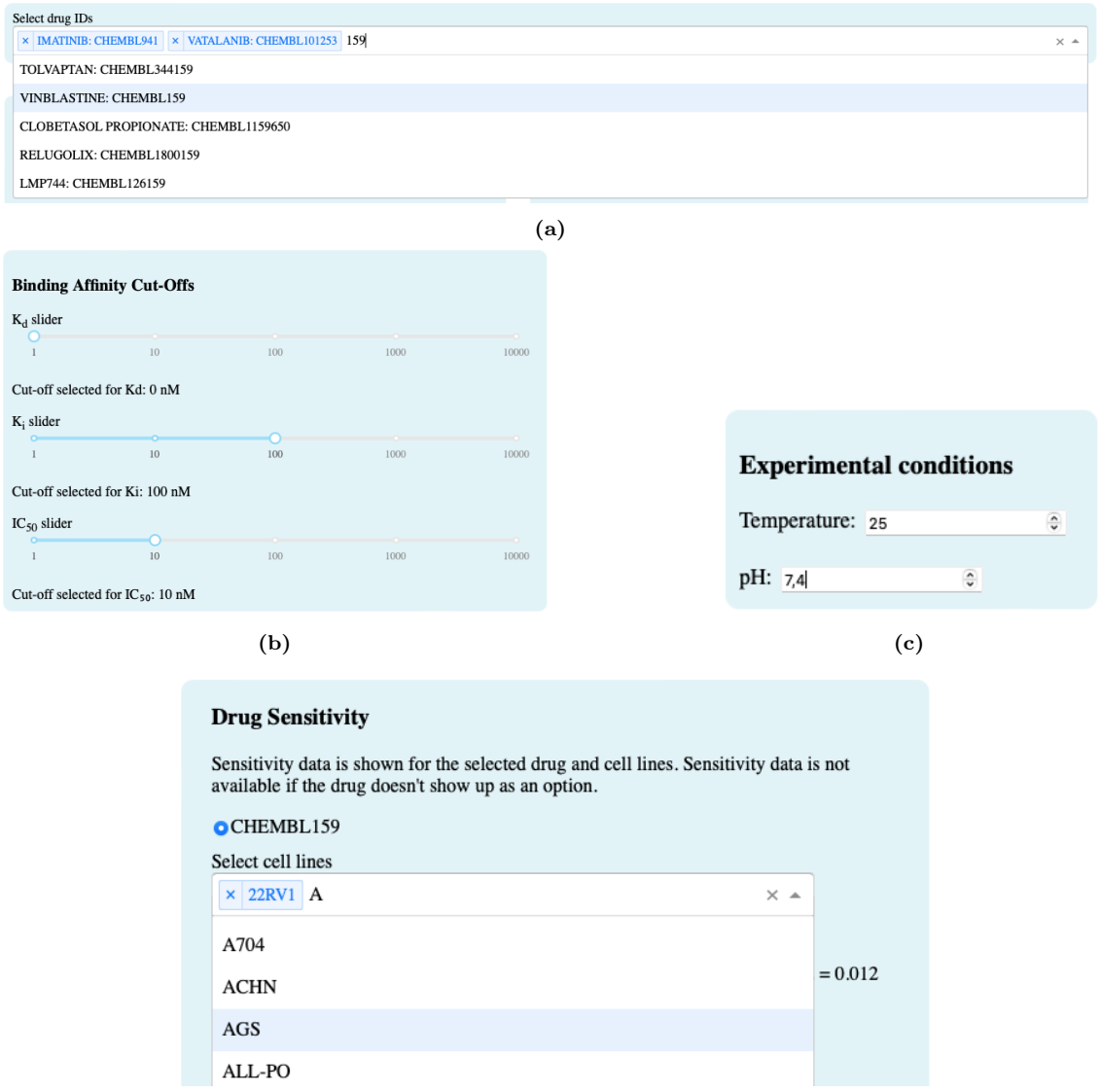


Figure 3.3: Input options of DrugProfiler tool **a)** Drop down menu for selection of drugs. **b)** Sliders for adjusting the binding affinity cut-offs and select measurement method (K_d, K_i, and IC_{50}). **c)** Input boxes to specify experimental conditions for binding affinity measurements. **d)** Radio items and dropdown menu to select drug and cell line, respectively, for sensitivity data.

3.2.2 Output Generated by the DrugProfiler Tool

The query results are presented as text and figures to facilitate the exploration of the identified target profiles. The different elements are shown in Figure 3.4.

Each profile is given in text format in the Drug Panel container (Figure 3.4a). Moreover, the profiles can be downloaded as a .txt file by clicking the "download-drug-panel". The user specifies the name of the file. The format corresponds to the drug panel input file used in the DrugLogics pipeline (3.4b). As a result, automated drug target profile annotation is facilitated, and the downloaded file can be used directly as input in the pipeline. However, direct utilisation requires that an HGNC annotation is applied when naming the target nodes.

As mentioned, target profiles are generated by identifying instances where the binding affinity value of a DTI falls below a certain threshold. Even if multiple binding affinity values are recorded for a DTI, only one value below the threshold is required to classify the molecule as a target. An alternative approach to classification could be using the median, which considers the distribution of binding affinity values for a given target. This approach could potentially improve the precision of the classification method, as it reduces the probability of misclassifying targets due to outlier values falling below a single cut-off threshold. In other words, fewer FP will probably be predicted. However, using the median may come at the expense of sensitivity, as some binding affinity values would be discarded, potentially leading to more FN. Additionally, using a single cut-off threshold offers more flexibility to adjust the threshold according to specific criteria. For instance, if the drug sensitivity for a cell line is measured at 1 µM, the cut-off could be set just above this limit to capture all potential targets. Finally, using the median could introduce bias if the distribution of binding affinity values is asymmetrical or if only a few values are recorded for the DT-pair. For example, if only two binding affinity values equal to 1 nM and 100 nM are stored for the pair, the median would be 50.5 nM, which may not accurately reflect the most frequent binding affinity strength of the target. To summarise, the minimum value was selected as the threshold to generate profiles that capture all potential targets. However, the distribution of binding affinity values should be studied to evaluate the confidence of target classification.

The figure and text in the Drug Target Network container facilitate a detailed analysis of each target profile. The inhibition profile is illustrated as a network with an edge between the drug node and each target node, representing the interactions. The figure has a zoom and drags view, making it easier to explore the profiles of numerous targets. The size of the target nodes is customised according to affinity strength. Larger nodes indicate stronger affinity (i.e. lower binding affinity value). The node colour indicates if one or multiple binding affinity measurements are provided below the set threshold, illustrated by light and dark grey, respectively. The target nodes can be clicked for detailed information about the DTI. The annotation defines the min and max value for binding affinity below the set cut-off and the standard error of the mean (Figure 3.4c). Hence, the variability across the recorded values can be examined to evaluate the confidence in the classification of each target.

Identifying DTI with considerable binding affinity is valuable when prioritising which target effects should be included when predicting the drug response. Furthermore, the information is vital to know which components should be present when selecting a model, modifying an existing model or building a model from scratch to capture the drug response better.

The Shared Target container facilitates the comparison of inhibition profiles generated for different drugs. The drugs selected for comparison are illustrated as nodes in a network with edges to the shared targets. The mutual targets are also listed in the text below the figure (Figure 3.4d). Analysing the similarity between target profiles when prioritising candidate drugs is beneficial. For

example, shared targets could indicate potential overlap in the drugs' target binding site. Consequently, the targets are likely to be functionally and structurally alike, increasing the probability of them being included in the same signalling pathway [71]. If the objective is to select drugs that target multiple components within one signalling pathway, shared targets are considered positive. On the contrary, shared targets might not be desired if the objective is to target distinct pathways.

Finally, the tool includes functionality to explore drug sensitivity data across cell lines. If sensitivity data is available for the selected drugs, they will show up as radio items. After specifying a drug, the cell lines of interest can be selected from a drop-down menu. The half maximum inhibitory concentration (IC_{50}) measured for the drug and the cell lines is described in the text (Figure 3.4e). As discussed earlier, this quantitative data is helpful when prioritising drugs to be applied in research performed on a specific cell line, for instance, to select drugs that require low dosages. Nevertheless, the parameter can be applied when defining the binding affinity cut-off. The cut-offs can be adjusted to match the given concentration to identify targets interacting with the drug below this threshold.

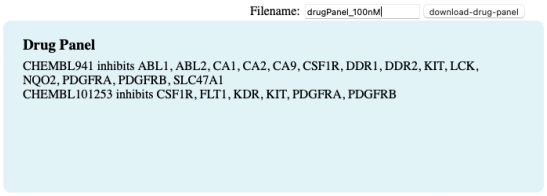
3.3 Drug Target Profiling

After finalising the Drug-Target Interaction database and the associated DrugProfiler tool, the new module was applied to generate target profiles for drugs of interest. The objective was to use the profiles to investigate how off-target effects influence drug response prediction by logical modelling. The 17 drugs selected for analysis have been studied by the DrugLogics initiative *in silico* and experimentally [29, 58, 72, 30]. The produced target profiles are presented in Table 3.2.

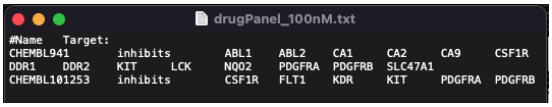
3.3.1 Definition of Query Conditions

Different cut-off values were applied to generate the target profiles because a dichotomy between high and low binding affinity is not universally defined, as multiple factors can influence affinity strength. However, binding affinity below 100 nM is often characterised as high affinity, between 100 nM to 10 µM as moderate, while binding affinity above 10 µM is considered weak [73, 59]. Cut-off values were increased progressively across the range of moderate to strong interactions to prioritise DTI with binding affinity in this range. Additionally, it was of interest to investigate whether considering more potential off-target effects improved prediction performance compared to only accounting for stronger interactions. The upper cut-off was set to 10 µM to exclude weaker binders, as these most likely do not bring about an effect. Finally, all three methods for measuring binding affinity (K_d, K_i, and IC_{50}) were included, and the same cut-off values were used for all to account for every potential off-target effect.

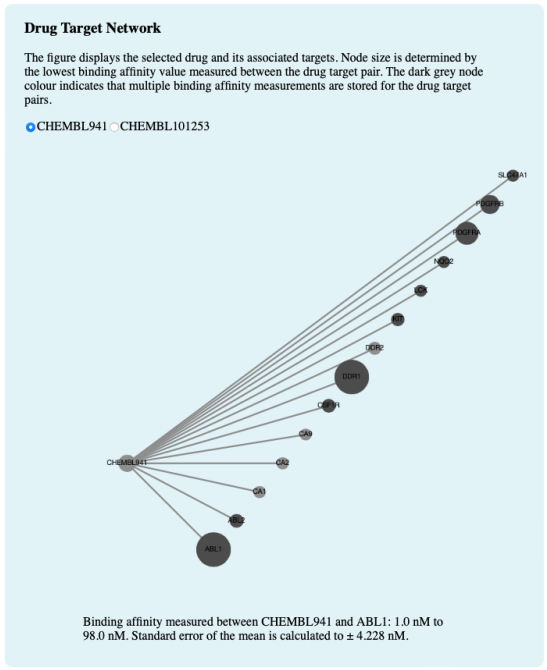
The experimental conditions for the binding affinity measurements were not specified when generating the target profiles. If pH and temperature are accounted for, all measurements must be conducted under the specified conditions, and NULL values are not permissible. Specifying these values would considerably restrict the scope of the analysis due to limited data availability for these attributes and variations in conditions across measurements. However, this is a weak point with the analysis because IC_{50} values depend on pH and temperature. Hence, a lack of specification may compromise the agreement across the affinity measurements.



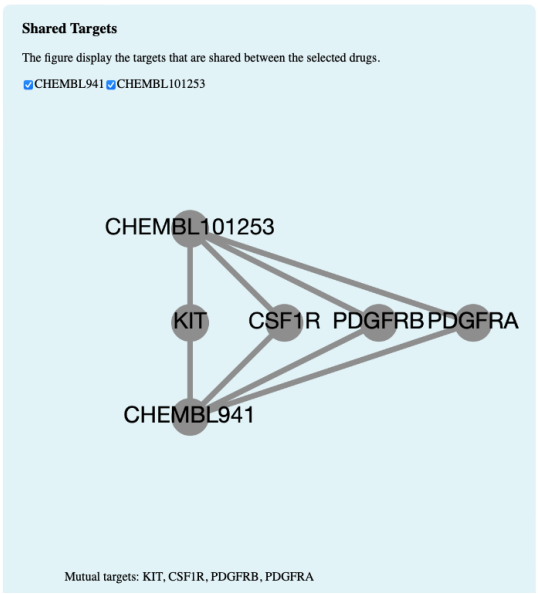
(a)



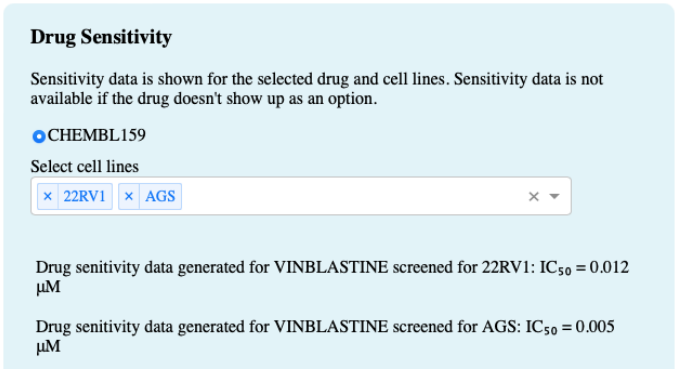
(b)



(c)



(d)



(e)

Figure 3.4: Output generated from queries carried out by the DrugProfiler tool. a) Drug panel representing the identified target profiles. b) Downloaded text file, where the set-up corresponds to the drug panel input file used in the DrugLogics pipeline. c) Network figure visualising a selected target profile. The size of each target node indicate the binding affinity strength. A low binding affinity value, indicating strong binding, is illustrated by a larger node. A darker node colour indicate that multiple measurements are stored for the drug-target interaction. Text stating the interval for binding affinity values and the standard error of the mean is generated when clicking a target node. d) Network figure and text allowing for comparison of target profiles by identifying the shared targets. e) Drug sensitivity data for a drug and selected cell lines stated by the half maximum inhibitory concentration (IC_{50}).

3.3.2 Evaluation of Drug Selectivity

Before the generated target profiles were used to predict drug effects, an exhaustive analysis of the profiles was conducted. The appearance of off-target effects, drug selectivity, and composition of targets were examined.

When evaluating the selectivity and appearance of off-targets across the drugs, a great span in the number of identified targets was observed. The target profiles generated with a 100 nM cut-off show that most of the analysed drugs show high affinity for one to three proteins, which tend to be primary targets. However, five drugs have five or more targets - namely: BIRB0796, PI103, BI-D1870, Ruxolitinib, and PRT 062607. When interpreting their profiles generated using the higher cut-offs, it is observed that the number of targets increases considerably. In comparison, the number of targets for the remaining drugs increases with a maximum of three. The only exception is PKF118-310. High affinity towards numerous additional targets indicates that these drugs have low selectivity and high off-target activity.

For example, CT99021 shows high specificity towards its primary targets, GSK3A and GSK3B. Only one off-target, CDK1, was identified, and the highest threshold was then applied. When comparing the results with the RSK inhibitor BI-D1870, similar results are observed at the lowest cut-off, which exclusively consists of the two primary targets, RSP6KA1 and RSP6KA3. When the cut-off is raised to 100 nM, the two remaining primary targets, RPS6KA2 and RSP6KA6, are included. However, strong affinity is also observed towards the off-targets PLK1 and VRK1. The number of off-target effects increases rapidly when the threshold is raised further, and nine additional entities are identified at the highest threshold. The large number of off-targets indicates that this inhibitor has much lower selectivity.

Molecular targeted therapies are designed to selectively modulate a single disease-causing target [74]. However, this approach to drug design has turned out to be challenging. The results show that most of the analysed drugs exhibit affinity towards additional targets to some extent. These findings agree with the significant obstacle of selectivity among small-molecule inhibitors, which is highly addressed [73, 75, 76]. The challenge is pointed out to be particularly widespread among kinase inhibitors. All five drugs, observed to show affinity towards numerous targets, are indeed kinase inhibitors.

Table 3.2: Drug name, ChEMBL ID, the primary target and target profiles for 17 chemical inhibitors. The target profiles were generated by the DrugProfiler tool, where the thresholds for K_d, K_i, and IC_{50} were set equal to 10 nM, 100 nM, 1 µM, and 10 µM.

Chemical Inhibitor	10 nM	100 nM	1 µM	10 µM
(5Z)-7-oxozeaenol	MAP2K1, MAP3K7	MAP2K1, MAP3K7, PLAA	MAP2K1, MAP3K7, MAPK1, PLAA	MAP2K1, MAP2K7, MAP3K7, MAPK1, PLAA
AKTi-1,2	AKT2	AKT1, AKT2	AKT1, AKT2	AKT1, AKT2, AKT3

Table 3.2 – continued from previous page

Chemical Inhibitor	10 nM	100 nM	1 µM	10 µM
BIRB0796	DDR1, DDR2, HTR2A, MAPK11, MAPK12, MAPK14, MAPK9, STK10, TIE1, TNIK	ABL1, BRAF, DDR1, DDR2, EPHA4, EPHA5, HSPB1, HTR2A, MAP4K4, MAPK10, MAPK11, MAPK12, MAPK13, MAPK14, MAPK9, MINK1, NTRK2, RET, SLK, STK10, TEK, TIE1, TNIK	ABL1, ANKK1, BLK, BRAF, CDC42BPB, CDK19, CDK5, CDK8, DDR1, DDR2, EPHA1, EPHA2, EPHA3, EPHA4, EPHA5, EPHA6, EPHA7, EPHA8, EPHB1, EPHB2, EPHB4, FLT1, FLT3, FLT4, FRK, HSPB1, HTR2A, KDR, KIT, LYN, MAP3K20, MAP3K7, MAP4K4, MAPK10, MAPK11, MAPK12, MAPK13, MAPK14, MAPK8, MAPK9, MINK1, NLK, NTRK1, NTRK2, NTRK3, PDGFRB, PTK2B, RET, SLK, STK10, TEK, TIE1, TNIK	ABL1, ABL2, ANKK1, BLK, BRAF, CDC42BPA, CDC42BPB, CDC42BPG, CDK13, CDK19, CDK5, CDK7, CDK8, CIT, CSF1R, DDR1, DDR2, EGFR, EPHA1, EPHA2, EPHA3, EPHA4, EPHA5, EPHA6, EPHA7, EPHA8, EPHB1, EPHB2, EPHB3, EPHB4, FGFR1, FGFR4, FGR, FLT1, FLT3, FLT4, FRK, HCK, HSPB1, HTR2A, KDR, KIT, LCK, LTK, LYN, MAP3K20, MAP3K7, MAP4K4, MAPK1, MAPK10, MAPK11, MAPK12, MAPK13, MAPK14, MAPK15, MAPK8, MAPK9, MERTK, MINK1, MKNK1, MKNK2, MUSK, NLK, NTRK1, NTRK2, NTRK3, PDGFRA, PDGFRB, PRKCB, PRKCD, PTK2, PTK2B, RAF1, RET, RIPK2, SLK, STK10, TEK, TIE1, TNIK, TTK, YES1
CT99021	GSK3A, GSK3B	GSK3A, GSK3B	GSK3A, GSK3B	CDK1, GSK3A, GSK3B
PD0325901	BRAF, MAP2K1	BRAF, MAP2K1, MAPK1	BRAF, MAP2K1, MAP2K2, MAPK1	BRAF, MAP2K1, MAP2K2, MAPK1
PI103	MTOR, PIK3C2B, PIK3CA, PIK3CB, PIK3CD, PIK3CG, PIK3R1, PRKDC	MAPKAP1, MLST8, MTOR, PIK3C2B, PIK3C2G, PIK3CA, PIK3CB, PIK3CD, PIK3CG, PIK3R1, PRKDC	DAPK3, HIPK2, HIPK3, MAP3K19, MAPKAP1, MLST8, MTOR, PIK3C2A, PIK3C2B, PIK3C2G, PIK3CA, PIK3CB, PIK3CD, PIK3CG, PIK3R1, PIP4K2C, PRKDC, RIOK2	BRAF, CDC42BPA, CDK7, CHUK, CLK1, CLK2, CLK3, CSNK2A1, CSNK2A2, DAPK1, DAPK2, DAPK3, DYRK1B, DYRK2, FLT3, HIPK1, HIPK2, HIPK3, JAK1, LIMK2, MAP3K19, MAPKAP1, MLST8, MTOR, MYLK, PI4KB, PIK3C2A, PIK3C2B, PIK3C2G, PIK3CA, PIK3CB, PIK3CD, PIK3CG, PIK3R1, PIP4K2C, PRKDC, RAF1, RIOK2, TYK2
PKF118-310	-	CTSB	ABL1, CDC25B, CTSB, CTSL, CTSS, DUSP1, DUSP3, HSP90AA1, KDM4C, NR5A1, SIRT1, TCF4	ABL1, CDC25B, CTNNB1, CTSB, CTSL, CTSS, DUSP1, DUSP3, HSP90AA1, KDM4A, KDM4C, NR5A1, PTPN22, PTPN7, SIRT1, TCF4, TCF7L2
JNK-IN-8	MAPK8, MAPK9, MAPK10	MAPK8, MAPK9, MAPK10	MAPK8, MAPK9, MAPK10	MAPK8, MAPK9, MAPK10

Table 3.2 – continued from previous page

Chemical Inhibitor	10 nM	100 nM	1 µM	10 µM
BI-D1870	RPS6KA1, RPS6KA3	PLK1, RPS6KA1, RPS6KA2, RPS6KA3, RPS6KA6, VRK1	AURKB, CSNK1A1, JAK2, PLK1, RPS6KA1, RPS6KA2, RPS6KA3, RPS6KA6, STK3, VRK1	AURKB, BRD4, CSNK1A1, GSK3B, JAK2, MARK3, PLK1, RPS6KA1, RPS6KA2, RPS6KA3, RPS6KA6, STK3, VRK1
BI605906	-	-	IKBKB	IGF1, IKBKB
SB-505124	-	ACVR1B, MAPK14, TGFBR1	ACVR1B, MAPK14, RIPK2, TGFBR1	ACVR1B, MAPK14, RIPK2, TGFBR1
Ruxolitinib	JAK1, JAK2, JAK3, TYK2	CAMK2A, CAMK2D, CAMK2G, DAPK1, DAPK2, DAPK3, DCLK1, GAK, HDAC6, JAK1, JAK2, JAK3, LRRK2, MAP3K2, ROCK1, ROCK2, TYK2	AAK1, ANKK1, BMP2K, BMPR2, CAMK1, CAMK1D, CAMK2A, CAMK2B, CAMK2D, CAMK2G, CLK2, DAPK1, DAPK2, DAPK3, DCLK1, DCLK2, DCLK3, DYRK1A, GAK, GRK1, GRK7, HDAC6, HIPK2, IRAK1, JAK1, JAK2, JAK3, LRRK2, LTK, MAP2K3, MAP3K19, MAP3K2, MAP3K3, MAP3K7, MARK2, MAST1, MKNK2, MYLK, NTRK2, NTRK3, NUAK2, PHKG2, PLK1, PLK3, PLK4, PRKCE, PRKG2, RET, ROCK1, ROCK2, RPS6KA1, RPS6KA2, RPS6KA4, RPS6KA5, RPS6KA6, STK16, TAOK2, TAOK3, TTK, TYK2, ULK1, ULK2	AAK1, ABL1, ALK, ANKK1, AURKC, BMP2K, BMPR1B, BMPR2, BRAF, CAMK1, CAMK1D, CAMK1G, CAMK2A, CAMK2B, CAMK2D, CAMK2G, CASK, CDK7, CLK2, CLK4, CSNK1A1, CSNK2A1, CSNK2A2, DAPK1, DAPK2, DAPK3, DCLK1, DCLK2, DCLK3, DMPK, DYRK1A, DYRK2, EPHA3, EPHB6, ERN1, GAK, GRK1, GRK4, GRK7, HDAC6, HIPK1, HIPK2, HIPK3, IKBKB, IKBKE, IRAK1, JAK1, JAK2, JAK3, KIT, LRRK2, LTK, MAP2K1, MAP2K2, MAP2K3, MAP2K4, MAP3K1, MAP3K15, MAP3K19, MAP3K2, MAP3K3, MAP3K7, MAP4K2, MAPK8, MARK2, MAST1, MELK, MKNK2, MYLK, NEK3, NEK7, NTRK1, NTRK2, NTRK3, NUAK2, OXSR1, PAK4, PAK5, PAK6, PHKG1, PHKG2, PI4KB, PLK1, PLK2, PLK3, PLK4, PRKCE, PRKG2, RET, RIOK1, RIOK2, RIOK3, RIPK4, ROCK1, ROCK2, RPS6KA1, RPS6KA2, RPS6KA4, RPS6KA5, RPS6KA6, SBK1, SBK3, SRPK1, SRPK3, STK16, STK17A, STK17B, STK25, STK26, STK39, TAOK1, TAOK2, TAOK3, TBK1, TNK2, TTK, TYK2, ULK1, ULK2, ULK3, VRK2
D4476	CSNK1D, CSNK1E	CSNK1D, CSNK1E	CSNK1A1, CSNK1D, CSNK1E, MAPK14	CSNK1A1, CSNK1D, CSNK1E, MAPK14, PRKD1
10058-F4	-	-	-	MYC

Table 3.2 – continued from previous page				
Chemical Inhibitor	**10 nM**	**100 nM**	**1 µM**	**10 µM**
Stattic	-	HIF1A	HIF1A	HIF1A, STAT3
GSK2334470	PDPK1	PDPK1	PDPK1	AURKB, PDPK1, ROCK1
PRT062607	SYK	FGR, MAP3K9, PRKCA, SYK, ZAP70	FGR, FLT3, LCK, LYN, MAP3K9, PAK5, PRKCA, SRC, SYK, YES1, ZAP70	FGR, FLT3, JAK2, KCNH2, LCK, LYN, MAP3K9, PAK5, PRKCA, SRC, SYK, YES1, ZAP70
End of Table				

3.3.3 Study of the Mechanism of Inhibition

After observing that several of the analysed drugs displayed affinity toward numerous targets, it was of interest to explore if the mechanism of inhibition influences the selectivity. It was, therefore, investigated how the DT binding occurs for the inhibitors that were observed to show a low degree of selectivity. The analysis was carried out by a literature study. When studying the mechanism of inhibition for the drugs that appeared to be less selective, all of them were found to be ATP-competitive inhibitors, except for BIRB0796.

PI103 is designed to be selective against the lipid kinase family of phosphatidylinositol 3-kinases (PI3K) and the mammalian target of rapamycin (mTOR). The drug has shown antitumour activity in cancers where the PI3K pathway is deregulated. PI103 binds to and forms H-bonds with the hydrophobic region I and the adenine pocket in the active ATP-binding pocket [77, 78]. BI-D1870 is an inhibitor of the p90 ribosomal S6 kinases (RSK). The RSK isoforms regulate the Ras-ERK1/2 signalling pathway. A pathway that often is altered in cancer. BI-D1870 inhibits its targets by binding to the N-terminal of the ATP-binding pocket [79]. Furthermore, Ruxolitinib is developed to inhibit Janus kinases 1 and 2 (JAK1/2). The drug works by disrupting the JAK / STAT signalling pathway. This compound binds to the ATP-binding pocket by hydrogen bonds to the hinge region and the catalytic site [80]. The last ATP-competitive inhibitor, PRT062607, targets the spleen tyrosine kinase (SYK) and alters the SYK/BCR signalling pathway. However, details about how the compound binds to the ATP binding pocket was unattainable.

BIRB0796 inhibits the stress-activated protein kinase p38 isoforms from the mitogen-activated protein kinase (MAPK) family. Unlike the other multi-targeted inhibitors, BIRB0796 interacts with an allosteric binding pocket and indirectly inhibits the binding of ATP [81]. The drug's inhibitory activity is achieved by interacting with the conserved Asp-Phe-Gly (DFG) motif at the beginning of the activation loop. The activation loop undergoes conformational reorganisations before binding ATP, transforming the kinase from inactive to active. This reorganisation makes the DFG motif accessible for substrate binding [82]. As the ATP binding pocket, the DFG motif is highly conserved across the kinome [83], which could contribute to the weakened selectivity of this inhibitor.

Characterisation of protein kinases has shown an extensive conservation in sequence and structure across the kinome. The highest degree of conservation occurs at the ATP binding pocket and the nearby regions [23]. As a result, it is incredibly challenging to design selective inhibitors that target this region. With this in mind, a competitive mechanism of inhibition was considered a highly possible reason for the reduced selectivity observed among these five drugs. Along with reduced

selectivity, additional challenges arise from this drug design. Firstly, the intercellular concentration of ATP can be raised by 5 mM, which is frequently observed in tumour cells. The concentration of ATP for most kinases to achieve half V_{max} (K_m) is in the µM range, and the raised ATP concentration would therefore lead to full-time saturation by ATP. Consequently, affinity in the nanomolar range is acquired to compete under these conditions. Secondly, numerous other ATP-dependent enzymes exist that are not kinases, which are essential to normal physiology. Drugs that mimic the ATP molecule to exhibit competitive inhibition could therefore lead to toxicity [76].

Despite these challenges, successful inhibitor design can be achieved using this ATP competitive mechanism of inhibition by binding to a more variable region [23]. Successful design is demonstrated by the results obtained from (5Z)-7-oxozeanol, which appears to be more selective than the drugs described above. The drug is a fungal resorcyclic acid lactone, which effectively inhibits mitogen-activated protein kinases. The aromatic ring of the resorcyclic lactone resembles the adenine ring of ATP, making the compound able to bind the ATP pocket as a competitive, irreversible inhibitor. However, this inhibitor requires the $Z_{5,6}$ keto-7enone functional group that functions as a Michael acceptor for a conserved cysteine residue found in the ATP-binding pocket in only a small group of the kinome [84]. This requirement is likely the cause of higher specificity towards a smaller subgroup, including TAK1, the MEK family and the ERK family [85]. MAP3K7 (TAK1), MAP2K1 (MEK1), MAP2K7, and MAPK1 (ERK2) are all observed in the generated target profile.

An alternative drug design is allosteric inhibition, where the agent does not compete with ATP. Development of such inhibitors is considered more difficult due to the challenge of identifying ligands for distinct regions of the protein kinases. However, they pose highly advantageous properties compared to ATP competitive inhibitors as regions outside the ATP binding pocket tend to be less conserved, facilitating a higher selectivity. Furthermore, because these inhibitors do not compete with ATP, concentrations closer to their biochemical K_i can be applied [86]. The result shows that AKTi-1,2 (AKT inhibitor VIII) appears highly selective. This drug is an allosteric and non-ATP competitive inhibitor which suppresses the phosphatidylinositol 3 (PI3) kinase/Akt pathway. The inhibitor binds to the N-terminal pleckstrin homology (PH) domain found in all three isoforms of human AKT. While strong affinity is observed towards AKT1 and AKT2, AKT3 is only included when the highest cut-off is applied. When comparing the proteins, AKT3 shows 83.6% and 78% identity with AKT1 and AKT2, respectively. However, the targeted PH domains show approximately 60% identity [87]. The results indicate that the inhibitor is rather selective towards isoforms of this domain.

3.3.4 Kinase Mapping

The final question to be investigated through profile analysis was whether the kinase targets within a profile belonged to the same kinase subfamily or if they were spread across the kinome. To examine how the drugs interact across the kinome, a phylogenetic tree of the human kinome was generated for each kinase inhibitor and annotated with the target profile. Each target was marked with the kinase name, and the annotation size was scaled according to the measured binding affinity. Kinase maps for all of the studied kinase inhibitors are shown in Figure A.1, in Appendix A.3 - Kinase Maps on page 76.

The figure shows that the specificity towards kinase subfamilies differs considerably among the analysed inhibitors. When examining the kinome trees generated for the drugs identified to be multi-targeted, binding affinity in the nanomolar range was observed towards targets spread across numerous groups. For instance, binding affinity is measured below 10 nM not only between BIRB0796 and the primary target MAPK14 (p38a) but also TIE1 and TNIK (Figure 3.5a). The three targets are members of three distinct kinase subfamilies, namely: CMGC, TK and STE. Such results

indicate a requirement for optimisation of this inhibitor's potency towards the primary target to enable the inhibitor to distinguish between the target of interest and off-targets.

Affinity towards multiple unrelated kinases is expected for ATP competitive kinases due to the evolutionary conservation of the substrate binding pocket across the kinome, discussed in the previous section. However, the selectivity of new kinase inhibitors is often evaluated through screening against a panel restricted to closely related kinases of the intended target. The idea is that interactions towards off-targets are more likely to occur within the subfamily with a more similar amino acid sequence [88]. The obtained results are proof of the contrary, where targets of the same inhibitor are members of multiple different kinase subfamilies, indicating that off-target effects are not exclusively revealed from the protein sequence. Consequently, it is insufficient to only screen for DTI within the subfamily of the intended target. The result highlights the requirement of extensive binding assays, including all kinase subfamilies, for adequate evaluation of selectivity and complete target profiling. The importance of this analysis should be emphasised to achieve successful development of new drugs and to achieve accurate predictions of drug effect.

More promising results were achieved for the drugs appearing to be more selective. Most of the targets identified for these agents were members of the same subfamily as the intended target. When targets were members of other subfamilies, the binding affinity for the interaction tended to be weaker. For instance, the complete set of targets identified for CT99021 is included in the CMGC family (Figure 3.5b). The second observation is illustrated by D4476 (Figure 3.5c). Most of this drug's targets are included in the CK1 family, and weak binding affinity is measured towards the two off-targets, MAPK14 (p38a) and PRKD1 (PKD1), which are members of the CMGC and CAMK families, respectively.

Such inhibitors, targeting a small number of kinases with high selectivity towards one subfamily, are promising research tools. Selective reagents are valuable aids allowing for analysis of the function of the targeted protein kinase. An acute need for such inhibitors is emerging from cell-signalling research, addressed by the number of citations on articles that introduce selective reagents [89].

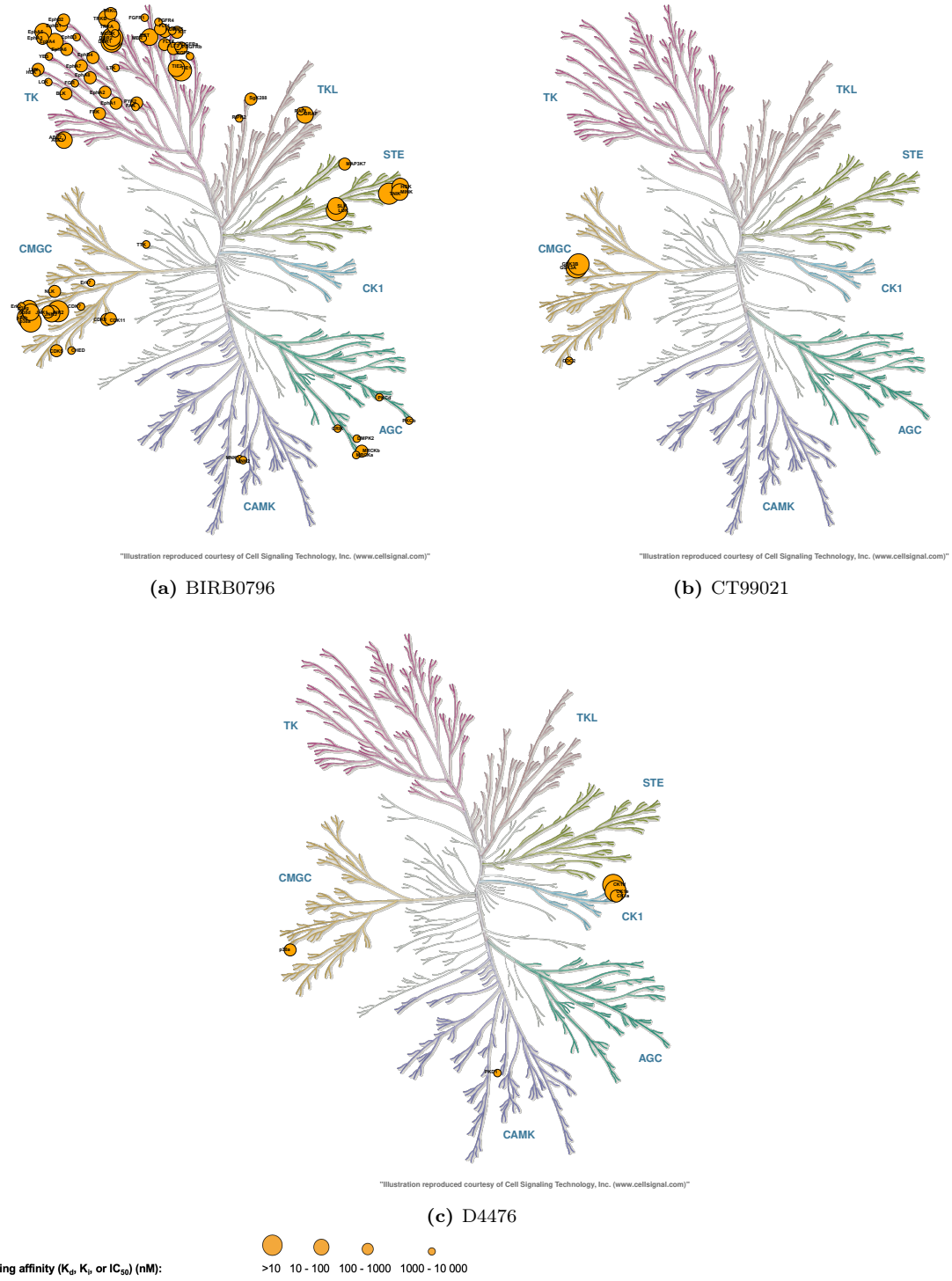
3.4 *In Silico* Prediction of Drug Effect

Using *in silico* simulations to predict drug effect is an important step forward in drug development and precision medicine. *In silico* simulations can be time and cost-effective. Even more important, the simulations can give insight into the mechanism of action and uncover potential side effects. A primary objective of this thesis was to investigate if accounting for off-target effects could improve predictions of drug effects and provide a more comprehensive understanding of the mechanism of action.

The *in silico* simulations were executed using two logical models, namely CAncer Signaling CAusality DatabasE (CASCADE) 1.0 and CASCADE 2.0. The models were perturbed according to the target profiles generated by the DrugProfiler tool and analysed in the previous section, which include off-target effects. Simulations were conducted by stochastic modelling with MaBoSS and with the DrugLogics software pipeline to investigate how off-target effects alter the predicted drug repsonse. Finally, the prediction performance of the DrugLogic software was evaluated by comparing the predicted synergies with gold standard synergies identified by experimental testing.

3.4.1 Filtrated Target Profiles

The applied logical models are focused toward specific signalling pathways. The target profiles were therefore required to be modified to exclude targets not present in the model. A script was



(a) BIRB0796

(b) CT99021

(c) D4476

Figure 3.5: Phylogenetic trees of the human kinome annotated with the kinases name of the target profile components for the kinase inhibitors a) BIRB0796, b) CT99021, and c) D4476. The figures are generated using KinMap, and the circle size indicates if binding affinity strength between the drug target pair is measured $<$ 10 nM, 10 nM to 100 nM, 100 nM to 1 µM, or 1 µM to 10 µM.

developed to carry out the filtration automatically and translate the target names from HGNC symbols to node names used in the model. The filtrated profiles used in simulations of CASCADE 1.0 and CASCADE 2.0 are given in Table A.6 and A.7, respectively, provided in Appendix A.4 - Filtrated Target Profiles, on page 80. The percentage of drug targets in each profile that are present in CASCADE 1.0 and 2.0 are given in Table 3.3 and 3.4, respectively.

The calculations show that numerous off-targets are not present in the applied models. Furthermore, the percentage of a target profile included in the model varies considerably across the drugs, from zero to one hundred per cent. For 43% of the drugs, the maximum proportion of targets included in CASCADE 1.0 is limited to one-quarter. Consequently, only a subset of the targets is considered in the modelling. In CASCADE 2.0, the number of drugs with only one quarter of targets present is reduced to 24%. A reduction is expected as CASCADE 2.0 is a more comprehensive model, which includes additional signalling pathways. The highest percentage is obtained < 10 nM for both models. High coverage is expected at low cut-offs because most drugs exhibit strong affinity toward their intended target, and all primary targets are included in the model.

The consequence of missing target nodes is that the model will not reflect the influence of inhibiting such a protein that may have critical functionality both in the targeted cell and in the whole organism. To demonstrate, less than 25% of BIRB0796's targets are presented in the two models. Affinity on the nanomolar range is measured towards the receptor tyrosine kinases TEK, which is not included in either of the two models. The protein is an angiopoietin receptor and plays a central role in the signalling pathways that regulate endothelial cell proliferation, migration and survival during angiogenesis, a capability included among the hallmarks of cancer [90, 6]. Modulating this DTI could potentially disrupt pathological angiogenesis and consequently have a therapeutic effect, which the applied models will not capture [91].

The high frequency of missing target nodes represents an important limitation of this study. When few targets are present in the model, predicting the drug effect becomes challenging. Moreover, the quality of the predictions will likely fluctuate due to varying percentages of included target nodes across the drugs. Missing targets are to be expected because these models are simplifications of reality. Selecting a larger model or performing model improvement should be considered to improve the capability of predicting drug effects. Introducing target nodes strongly inhibited by a drug or with critical functionality in the analysed disease should be emphasised in such a process.

The challenge of reflecting the complexity of a biological system represents one of the main bottlenecks of mathematical modelling in biology. A model is, by nature, a simplified representation of reality, only illustrating parts of the original. Although a model can be expanded, it is essential to realise that not even an individual patient can represent a flawless model system for an entire disease as no patients are truly identical [92].

With this in mind, the statement "essentially, all models are wrong, but some are useful" by the statistician George Box should be considered [46]. Although all models represent only specific aspects of reality and always will be imperfect, they can still contribute valuable insight. A cut-off is often sufficient since the objective of the modelling is to answer a specific question. To conclude, the model must be correctly utilised, and the limitations must be recognised in the context of the problem to be solved [46].

Table 3.3: Drug name and percentage of components in the utilised target profiles included in the CASCADE 1.0 network model. The target profiles were generated by the DrugProfiler tool, using cut-offs for K_d, K_i, and IC_{50} equal to 10 nM, 100 nM, 1 µM, and 10 µM.

Chemical inhibitor	**10 nM**	**100 nM**	**1 µM**	**10 µM**
(5Z)-7-oxozeaenol	100%	67%	75%	80%
AKTi-1,2	100%	100%	100%	75%
BIRB0796	20%	13%	23%	23%
CT99021	100%	100%	100%	67%
PD0325901	100%	100%	100%	100%
PI103	25%	27%	17%	18%
PKF118-310	-	0%	8%	18%

Table 3.4: Drug name and percentage of components in the utilised target profiles included in the CASCADE 2.0 network model. The target profiles were generated by the DrugProfiler tool, using cut-offs for K_d, K_i, and IC_{50} equal to 10 nM, 100 nM, 1 µM, and 10 µM.

Chemical inhibitor	**10 nM**	**100 nM**	**1 µM**	**10 µM**
(5Z)-7-oxozeaenol	100%	67%	75%	80%
AKTi-1,2	100%	100%	100%	75%
BIRB0796	20%	13%	23%	24%
CT99021	100%	100%	100%	67%
PD0325901	100%	100%	100%	100%
PI103	25%	27%	17%	23%
PKF118-310	-	0%	8%	18%
JNK-IN-8	67%	67%	67%	67%
BI-D1870	100%	67%	60%	54%
BI605906	-	-	100%	50%
Ruxolitinib	50%	18%	19%	17%
SB-505124	-	100%	75%	75%
D4476	100%	100%	100%	80%
10058-F4	-	-	-	100%
Stattic	-	0%	0%	50%
GSK2334470	100%	100%	100%	67%
PRT062607	100%	40%	36%	38%

3.4.2 Predicted Cell Fate Phenotypes

To investigate how off-target effects may influence drug response, stochastic simulations with MaBoSS were performed. The influence was examined from alterations in the computed cell viability. Viability was calculated by subtracting the value of active *Antisurvival* output nodes from the value of *Prosurvival* output nodes. The degree of cell growth is ranged from -3 to +3, indicating total antisurvival and strong proliferation, respectively.

Modelling was first conducted using CASCADE 1.0, initiated by a simulation without any model perturbations. The results showed a 100% probability for prosurvival, indicating strong proliferation. Strong proliferation was expected because the unperturbed model is tailored to represent actively growing cancer cells.

Next, perturbations were introduced according to the target profiles in Table A.6 (Page 80). The results showed that the cell growth was predicted to be strongly proliferative when the model was perturbed by (5Z)-7-oxozeaenol, BIRB0796 or CT99021 when only the primary targets were considered. In contrast, decreased cell growth was predicted for the two first compounds when accounting for off-target effects. Altered cell growth is also predicted for PKF118-310 when off-target effects are considered. These results indicate that off-target effects tend to enhance the drug response. The predicted probabilities for the level of cell growth, obtained through modelling with CASCADE 1.0, are provided in Table 3.5.

(5Z)-7-oxozeaenol inhibits TAK1 (MAP3K7) and the off-target MEK1 (MAP2K1), with strong binding affinity (< 10 nM). The target proteins are included in the NF-kappa B and MAPK signalling pathways, respectively. The cell growth is predicted to be reduced from a 100% probability for level 3, when only accounting for the primary target, to level 2 or 1, with approximately 58% and 42% probability, respectively, when off-targets are included. The same result is observed for all four cut-offs. MEK/TAK1 was one of the drug synergies predicted and verified by Flobak et al. (2015) [29]. The combination of inhibition is also shown to facilitate apoptosis in KRAS-dependent cells, a dependency found in adenocarcinoma cancer cell line (AGS) [93, 94].

When predicting the effect of BIRB0796, cell growth was not observed to be reduced before off-target effects identified > 100 nM were considered. The target profile below the 100 nM cut-off includes p38alpha, JNK and Raf. All of these proteins are included in the MAPK signalling pathway. Reduced cell growth was predicted when the activity of NLK, TAK1, and RTPK was turned off. These proteins are included in other central pathways, namely: the WNT signalling pathway (NLK), signalling by RTK (RTPK) and the NF-kappa B signalling pathway (TAK1). All of these pathways are known to be deregulated in gastric adenocarcinoma.

Similar results are obtained from the predictions performed for PKF118-310. The calculated cell growth is reduced considerably, from 2 to -2, when TCF and beta-catenin from the WNT signalling pathway and DUSP1 from the MAPK signalling pathway are inhibited. This trend could imply that inhibiting multiple components in various pathways has a more significant effect than inhibiting one or more components in a single pathway.

The limited effect obtained when only one target or pathway is modulated is expected due to the complex, networked system of the cancer tumour. The disease arises from an accumulation of molecular alterations causing the deregulation of several signalling pathways. Moreover, crosstalk between signalling pathways makes the system redundant and adaptable. Crosstalk appears when entities from one signalling pathway bind, activate or inhibit entities that are members of distinct pathways [46]. Crosstalk results in alternative pathways that enable the tumour to cope with the loss of one or more components causing modifications of only one pathway [11]. In conclusion, the results suggest that successful drug effects may require modification of multiple pathways.

Multiple large-scale functional genomic projects performed in different model organisms demonstrate the challenge of accomplishing successful targeted therapy by modulating only one component [11]. The studies show that most independent single-gene knock-outs show little or no effect on the phenotype under laboratory conditions. In fact, no more than approximately 19% of genes are proven to be essential among a number of model organisms [95, 96, 97]. Furthermore, a study where each druggable gene in the mouse genome was knocked out, and each deletion was analysed across a panel of phenotypic assays revealed that only 10% of the deletions resulted in phenotypes showing potential for drug target validation [98].

The frequent observation of reduced cell growth, seen when off-target effects are regarded, shows the advantage of multi-targeted drugs in cancer therapy. The results indicate the need for a paradigm shift from the earlier dominating 'one-target-one-drug-one-disease' philosophy to polypharmacology or drug synergies. This requirement is highlighted in multiple articles [26, 99, 100].

The benefit of multi-targeted drugs is not as apparent from the results obtained through simulations with CASCADE 2.0. Ten additional drugs were analysed, and perturbations were introduced according to the target profiles in Table A.7 (Page 81). When interpreting the calculated probabilities for cell growth, shown in Table 3.6, an alteration is observed for a lower percentage of drugs, compared to the results obtained from simulations with CASCADE 1.0. Only five of the seventeen drugs showed reduced cell growth when including off-target effects. Moreover, when comparing the seven drugs simulated with the first model, the level of cell growth for (5Z)-7-oxozeaenol is no longer predicted to be altered while PI103 now show reduced cell growth.

When analysing the cell growth computed at the highest cut-off, almost half of the drugs show strong proliferation (Prosurvival = 3 and Antisurvial = 0). Hence, the network model appears less influenced by perturbations. Lack of altered cell growth indicates that CASCADE 2.0 is more robust, compared to CASCADE 1.0. This characteristic is anticipated due to the additional nodes and signalling pathways introduced in this model. Consequently, the complex network with alternative and compensatory signalling routes observed in cancer tumours is better captured.

Robustness is one of the key features of biological systems. The feature is defined as a system's capability to maintain biological function when exposed to perturbations. Robustness can be attained by transitioning from one steady state to another to tolerate perturbations and maintain biological function. This phenotypic switch is shown to be influential in, for instance, the development of drug resistance [28].

Network analysis indicates that the robustness observed in biological systems is rooted in the structure of biological pathways and interactions. Biological networks tend to be scale-free, where the node degree distribution follows a power-law, with a majority of nodes having a low node degree and a few hub nodes having a high degree. This characteristic makes biological systems resilient against the failure of random components because the probability of selecting a node with a large number of connectors is low. Eliminating such nodes will not change the path structure of the remaining nodes and will, therefore, not affect the overall topology of the network. On the other hand, a small set of nodes are highly connected hubs. Scale-free networks significantly depend on these components as they are essential to retaining the network's connectivity. When eliminating hubs, the network's diameter will increase considerably. The resulting topology will be significantly altered with a reduced communication capability between the nodes. This dependency makes the network extremely vulnerable to attacks on hubs [101].

When interpreting the obtained results in light of the scale-free network characteristic exhibited by the cancer system, it can be concluded that targeting multiple random nodes is ineffective. Even an attack on the disease-causing genes will not necessarily be a successful approach. Conversely, the topology of the network nodes should be emphasised to attack highly influential nodes.

Table 3.5: The predicted probability of cell growth, scaled from -3 (maximum antisurvival) to +3 (maximum prosurvival). The growth sum was calculated from network states identified by stochastic modelling on the logical model CASCADE 1.0 using the simulation software MaBoSS. Simulations were performed after perturbing the model according to target profiles generated for seven inhibitors, including the primary target, and off-target effects measured below the thresholds 10 nM, 100 nM, 1µM, and 10 µM

Chemical inhibitor	Primary target	10 nM	100 nM	1 µM	10 µM
(5Z)-7-oxozeaenol	100%: 3	57.5%: 2 42.5%: 1	58.3%: 2 41.7%: 1	57.7%: 2 42.3%: 1	58.0%: 2 42.0%: 1
AKTi-1,2	100%: 2	100%: 2	100%: 2	100%: 2	100%: 2
BIRB0796	100%: 3	100%: 3	100%: 3	50.6%: 3 49.4%: 2	100%: 2
CT99021	100%: 3	100%: 3	100%: 3	100%: 3	100%: 3
PD0325901	58.1%: 2 41.9%: 1	57.5%: 2 42.5%: 1	58.3%: 2 41.7%: 1	58.3%: 2 41.7%: 1	58.3%: 2 41.7%: 1
PI103	100%: 1	100%: 1	100%: 1	100%: 1	100%: 1
PKF118-310	100%: -1	-	-	100%: 2	100%: -2

Table 3.6: The predicted probability of cell growth, scaled from -3 (maximal antisurvival) to +3 (maximal prosurvival). The growth sum was calculated from network states identified by stochastic modelling on the logical model CASCADE 2.0 using the simulation software MaBoSS. Simulations are performed after perturbing the model according to target profiles generated for 17 inhibitors, including the primary target, and off-target effects measured below the thresholds 10 nM, 100 nM, 1µM, and 10 µM.

Chemical inhibitor	Primary target	10 nM	100 nM	1 µM	10 µM
(5Z)-7-oxozeaenol	100%: 3	100%: 3	100%: 3	100%: 3	100%: 3
AKTi-1,2	100%: 2	100%: 2	100%: 2	100%: 2	100%: 2
BIRB0796	100%: 3	100%: 3	100%: 3	59.7%: 2 40.3%: 1	50.1%: 1 49.9%: 3
CT99021	100%: 3	100%: 3	100%: 3	100%: 3	100%: 3
PD0325901	100%: 3	100%: 3	100%: 3	100%: 3	100%: 3
PI103	100%: 2	100%: 2	100%: 2	100%: 2	59.2%: 2 40.1%: 1 0.7%: Others
PKF118-310	100%: 3	-	-	100%: 2	100%: 2
JNK-IN-8	100%: 3	100%: 3	100%: 3	100%: 3	100%: 3
BI-D1870	100%: 2	100%: 2	100%: 2	100%: 2	100%: 2
BI605906	100%: 3	-	-	100%: 3	100%: 3
Ruxolitinib	100%: 3	100%: 3	100%: 3	100%: 2	100%: 2
SB-505124	100%: 3	-	100%: 3	100%: 3	100%: 3
D4476	100%: 3	100%: 3	100%: 3	100%: 3	100%: 3
10058-F4	100%: 2	-	-	-	100%: 2
Stattic	100%: 3	-	-	-	100%: 3
GSK2334470	100%: 2	100%: 2	100%: 2	100%: 2	100%: 2
PRT062607	100%: 3	100%: 3	100%: 3	71.3%: 2 28.1%: 1 0.64%: Others	71.7%: 2 27.4%: 1 0.48%: Others

3.4.3 Drug Synergies Prediction

Drug combinations that modify components included in different signalling pathways or cellular processes have been shown to enhance drug effect, similar to multi-targeted drugs. As a result, drug resistance development can be postponed, and drug dosage decreased to prevent drug-induced toxic effects. The DrugLogics pipeline is developed to run *in silico* simulations to identify such drug synergies. However, off-target effects have not been accounted for in their previous studies [29, 30]. As demonstrated, additional targets are frequently observed and tend to influence the drug effect. Including off-target effects is, therefore, hypothesised to enhance the software's ability to predict synergies. The objective of this section is to test this hypothesis.

Simulations with the pipeline were executed using CASCADE 1.0 and CASCADE 2.0, and the models were perturbed according to the target profiles in Table A.6 and A.7 (Page 80 and 81), respectively. The analysed combinations were prioritised according to the predicted synergy scores. Furthermore, each combination was classified as TP, FP, TN, or FN, using gold standard synergies found by experimental testing [29] [58]. The True Positive Rate (TPR) (sensitivity, recall), the False Positive Rate (FPR) and the precision were calculated from the classifications and used to plot receiver operating characteristic (ROC) and precision-recall (PR) curves. Prediction performance was interpreted from the AUC values obtained from these curves to evaluate if accounting for off-target effects leads to an improvement. Figure 3.6 shows an overview of the AUC values obtained from the various simulations. All ROC and PR curves generated for prediction performance are provided in Appendix A.5, on page 83.

When interpreting the synergy scores calculated for the TP combinations predicted through simulations with CASCADE 1.0, it is observed that nearly all are calculated to be more negative when off-target effects are considered compared to only the primary target. Table 3.7 shows the predicted synergy scores. A more negative score indicates that the predicted combinations show a stronger synergistic effect when off-targets are included.

The predicted ROC AUC values, presented in Figure 3.6a, suggest that accounting for off-target effects improves prediction performance when the models are trained to a steady state. The top four predictions include all the observed synergies when perturbing the model according to the target profiles generated using the 10 nM, 100 nM, and 10 µM cut-offs. Consequently, a ROC AUC value equal to one is obtained, meaning that the pipeline in these cases can accurately discriminate between TP and FP. A slight decrease is observed for the ROC AUC calculated for the 1 µM cut-off. The weakened result is caused by the FP combination consisting of AKTi-1,2 - PKF118–310, which was predicted to be the most synergistic combination in this simulation.

The three corresponding PR AUC values support the hypothesis that including off-target effects enhance prediction performance. The AUC value is once more calculated to be one. Hence, the pipeline achieves high precision suggesting a low FPR and a high recall suggesting a low false negative rate. To summarise, the output simulation results were highly accurate, indicating that the software's ability to predict synergistic combinations is enhanced when off-target effects are included. In theory, the accurate predictions decrease the experimental search space down to only true synergies, leading to more efficient and economical testing.

The ROC AUC and PR AUC calculated for the ensemble of models calibrated to a random proliferative phenotype are all lower or equal compared to the ensemble of models calibrated to AGS steady state. These results show the opposite of what is observed when only the primary target is considered, where normalisation of topology-intrinsic prediction propensities was necessary for improved prediction performance. Hence, the models trained to steady-state data perform considerably better than random model predictions.

The prediction performance is also observed to be improved when accounting for off-targets in simulations with the CASCADE 2.0 model. When interpreting synergy scores predicted for TP combinations, they are observed to be reduced when off-target effects are considered compared to only the primary target, corresponding to the result obtained from CASCADE 1.0. However, the ROC AUC values, presented in Figure 3.6c, show that there is an overlap between the TPs and the FPs. Additionally, when interpreting the synergy scores, shown in Table 3.8, it is seen that the simulation output for the three lowest cut-offs includes either two or three FN. Hence, the software is not able to perfectly distinguish between synergistic combinations and those that are not.

A better ROC AUC is obtained when off-target effects measured above 100 nM are taken into account, compared to only the primary target. Moreover, all of the observed synergies are predicted at the highest cut-off. The improved result obtained for the highest cut-off, for both CASCADE 1.0 and 2.0, implies that DTIs with moderate binding affinity influence the drug effect. The results suggest that such off-targets should be included when constructing the model to increase the probability of capturing all true synergies and improve the software's capability to distinguish between TP and FP.

The cut-offs for binding affinity were selected to allow for scaling the DTI strength and to evaluate if interactions with lower affinity are of value to improve prediction performance. However, simulations are not conducted with a model perturbed according to target profiles generated by applying the concentrations used during experimental testing of the synergies as binding affinity cut-offs, which is considered a limitation of this study. Such simulation can be executed to better model the actual experimental conditions, and it could be anticipated that these modifications would to lead to improved prediction performance.

The PR AUC values achieved when performing simulations with CASCADE 2.0 were considerably lower compared to simulations with CASCADE 1.0 (Figure 3.6d). Simulations in which the model is perturbed according to the primary target and target profiles generated for the three lower cut-offs did not capture all of the observed synergies (see Table 3.8). The produced FN leads to lower recall. Consequently, the experimental space could not be reduced without missing out on true synergies. Three drug combinations were predicted as FN, namely: BIRB0796 - BI-D1870, PI103 - GSK2334470, and AKTi-1,2 - BIRB0796. When interpreting the combinations, it is found that BIRB0796 and PI103 have notably low coverage of targets in the applied model (see Table 3.4). Hence, multiple off-target effects are not included in the simulations because nodes representing these targets are missing. The excluded effects cause a restricted reflection of reality and are a possible explanation for the weakened recall. The model should be considered modified to include additional drug targets to examine if accounting for more off-target effects could improve the performance.

Furthermore, numerous combinations predicted as synergies were not observed experimentally, resulting in a low precision (Figure A.5). Low coverage does not appear as the vital reason for predicting FP when evaluating the percentage of targets present for drugs included in the combinations with the lowest synergy score. A potential explanation for low precision can be errors in the prior knowledge networks (PKN). Flobak et al. (2021) introduced random modifications to the regulatory edges in this model, and their results indicated that even minor levels of randomisation in the network topology lead to notable errors in the predictive power. The PKN could be reevaluated to investigate if correcting prior knowledge could enhance the performance [72].

Although the PR AUC values obtained for CASCADE 2.0, in general, are relatively weak, a slight increase is observed when the model is perturbed according to target profiles generated at all four cut-offs.

To summarise, the results suggest that accounting for off-target effects is valuable in enhancing

prediction performance. In addition, the importance of including important off-targets when selecting target nodes during model construction is emphasised. It is expected, that modifying the applied models to incorporate knowledge of additional targets could enhance their adequacy for predicting drug synergies. Improved models would provide a more reliable basis for drawing a definitive conclusion.

Table 3.7: Drug combination and predicted synergy scores for experimentally validated drug combinations, obtained through simulations with CASCADE 1.0 using the DrugLogics pipeline (HSA synergy method, 150 models calibrated to the adenocarcinoma cancer cell line (AGS) steady state). The model was trained to steady state and perturbed with the primary target and according to target profiles generated by the DrugProfiler tool. The applied cut-offs for K_d, K_i, and IC_{50} were 10 nM, 100 nM, 1 µM, and 10 µM, respectively.

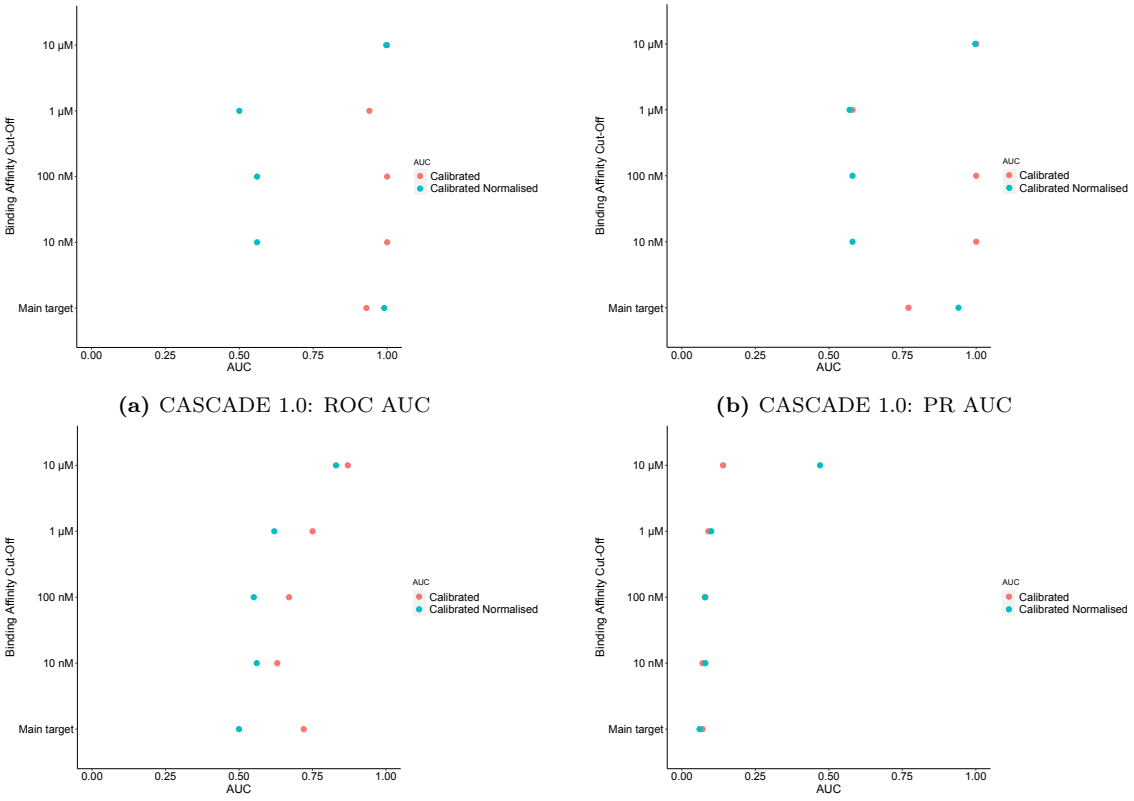
Perturbation	Primary Target	10 nM	100 nM	1 µM	10 µM
PD0325901 - AKTi-1,2	-0.94282	-0.94282	-1.23136	-1.23136	-1.23136
PI103 - PD0325901	-0.63369	-1.15821	-1.11820	-1.11820	-1.44836
AKTi-1,2 - (5Z)-7-oxozeaenol	-0.46254	-0.94282	-0.94282	-1.23136	-1.23136
PI103 - (5Z)-7-oxozeaenol	-0.31361	-1.15821	-1.15821	-1.11820	-1.44836

Table 3.8: Drug combination and predicted synergy scores for experimentally validated drug combinations, obtained through simulations with CASCADE 2.0 using the DrugLogics pipeline (HSA synergy method, 150 models calibrated to the adenocarcinoma cancer cell line (AGS) steady state). The model was trained to steady state and perturbed with the primary target and according to target profiles generated by the DrugProfiler tool. The applied cut-offs for K_d, K_i, and IC_{50} were 10 nM, 100 nM, 1 µM, and 10 µM, respectively.

Perturbation	Primary Target	10 nM	100 nM	1 µM	10 µM
PI103 - PD0325901	-0.25111	-0.47773	-0.44555	-0.44555	-0.68048
PI103 - (5Z)-7-oxozeaenol	-0.01834	-0.47773	-0.47773	-0.44555	-0.30243
BIRB0796 - BI-D1870	-0.02758	0	-0.01785	0	-0.91103
PI103 - GSK2334470	-0.02013	0	0	0	-0.10422
PI103 - BI-D1870	-0.02000	-0.03000	-0.03260	-0.03260	-1.00563
AKTi-1,2 - BIRB0796	0	0	0	-0.51515	-0.51929

3.5 Use of Drug Target Profiling in Selection of Cancer Treatment

The final objective in this thesis was to apply the DrugProfiler tool to investigate if a cancer patient could benefit from treatment with an off-label drug. The patient participates in the IMPRESS-Norway study, where off-label treatment is offered based on the patient's molecular profile [68]. The treatment was evaluated based on the target profile's overlap with the mutations identified from the gene panel analysis of the patient's tumour. The molecular profile showed gene amplification of KIT, KDR, and PDGFRA. The off-label drug of interest was *Imatinib*, and the target profile was generated using the DrugProfiler tool. KIT and PDFGRA were identified as drug targets, and the lowest binding affinity values recorded for the DTIs were 13 nM and 2 nM, respectively. For this reason, strong inhibition of the protein activity is expected at low dosages. Furthermore, multiple records were stored for each interaction, indicated by the dark grey node colour, increasing the confidence that these are TP targets. The findings imply that the patient could benefit from treatment with *Imatinib*. Figure 3.7 shows the target profile produced for the drug.



(a) CASCADE 1.0: ROC AUC

(b) CASCADE 1.0: PR AUC

(c) CASCADE 2.0: ROC AUC

(d) CASCADE 2.0: PR AUC

Figure 3.6: Plots showing area under the curve (AUC) values calculated for receiver operating characteristic (ROC) curves in the left panels and precision-recall (PR) curves in the right panels. The curves were created from the output obtained from simulations with (a) (b) the CASCADE 1.0 and (c) (d) the CASCADE 2.0 model (HSA synergy method, 150 models calibrated to the adenocarcinoma cancer cell line (AGS) steady state, represented by red dots, and normalised to a random proliferative phenotype, represented by blue dots). The models were perturbed using a) the primary target and target profiles generated 7 and 17 drugs, respectively. The profiles were generated with the DrugProfiler tool, using binding affinity cut-offs equal to b) 10 nM, c) 100 nM, d) 1 µM, and e) 10 µM.

The example shows how target profiling can facilitate precision medicine. The approach makes it possible to identify drugs with a high affinity toward one or several proteins. Moreover, the identified targets can be matched against the patient's molecular profile. A good fit can lead to greater effectiveness at lower dosages, hopefully resulting in reduced toxicity and postponed development of resistance. In addition, affinity towards targets that can cause side effects can be discovered.

Furthermore, the example demonstrates how target profiling can uncover potential for drug off-label use. High affinity towards additional targets can uncover new areas of application. *Imatinib* is an excellent example of this. The drug is a competitive ATP inhibitor, developed initially to treat patients with chronic myeloid leukaemia positive for the Philadelphia chromosome. The Philadelphia chromosome is a genetic abnormality caused by the translocation of chromosomes 9 and 22, and the result is an oncogene BCR-ABL1 gene fusion [102]. ABL1 is the primary target of *Imatinib*, bound with high affinity (Figure 3.7). However, the drug exhibits additional target effects, including inhibition of the two proteins KIT and PDGFRA. The multi-targeted property shows high potential for off-label use. Successful off-label use of *Imatinib* is demonstrated in, for instance, gastrointestinal stromal tumours frequently caused by a KIT or a PDGFRA mutation and in metastatic melanomas with KIT amplifications [103, 104].

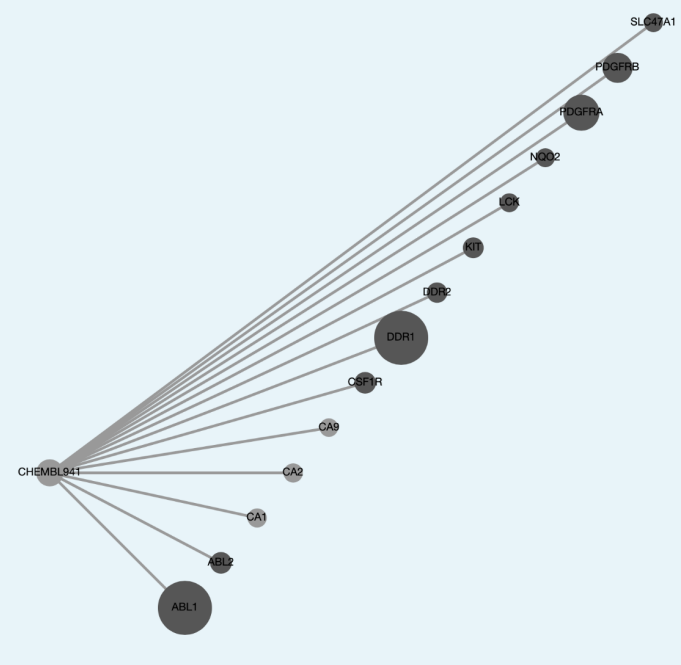


Figure 3.7: Network illustrating the target profile generated for *Imatinib* (CHEMBL941) using the DrugProfiler tool. The cut-off was set equal to 100 nM. The node size illustrates the strength of the binding affinity. The dark grey node colour indicates that multiple measurements are performed for the drug-target interaction (DTI) below the set cut-off.

4 Conclusion

The objective for this project was to investigate how model based predictions are altered when off-target effects are included in simulations. The first main aim was to develop a new module connected to the DrugLogics pipeline that facilitates automated drug target profiling, accounting for off-target effects. The second main aim was to use the new module to generate target profiles and apply these to study how prediction of drug reponse were changed when accounting for off-target effects. In this section, an evaluation was conducted to determine to what extent these purposes have been fulfilled, taking into consideration the key findings, limitations, and potential future prospects.

For the first objective, a database application was successfully designed and implemented, consisting of a relational database and associated programs implemented for updates and queries. The database design supports efficient storage of data on drugs, their targets and the prior knowledge available on drug-target interaction (DTI). Data from multiple data resources was homogenised and structured using the implemented programs before loaded into the application. The finalised database provides extensive information about drugs and their target space, allowing for comprehensive analysis of drug target profiles. However, data retrieval is required when new data is provided by the applied resources, to maintain an updated database.

The DrugProfiler tool makes the database accessible for the end user independent of programming background. The tool facilitates automated target profiling, where the output is customised according to user-specified conditions. Furthermore, analyses of the generated profiles are enabled by a Graphical User Interface (GUI) consisting of figures and text for exploration and assessment of the query outputs. A key functionality is the possibility to download drug panels that can be applied as input in the DrugLogics software.

For the second objective, target profiles were generated for 17 drugs using the new module, and these were thoroughly analysed. Off-target effects were identified for several drugs, indicating that they are multi-targeting. Investigation of the molecular mechanisms of inhibition implied that low selectivity could be a result of the target regions being highly conserved among kinase. Additionally, further analyses showed that drugs with high off-target activity showed low selectivity towards a single kinase subfamily.

The generated target profiles were used in *in silico* modelling to investigate how off-target effects altered the predicted drug response. Accounting for off-target effect was frequently predicted to enhance drug response, indicated by reduced cell viability. The most significant change was identified when targets were included in distinct signalling pathways. Furthermore, the analysis implied that more robust networks were less affected by random off-target effects.

Subsequently, the target profiles were applied in simulations with the DrugLogics pipeline. The obtained results indicated that accounting for off-target effects could lead to improved prediction performance. Best performance was achieved when targets bound with both moderate and high affinity were included in the simulation. Interpretation of combinations predicted to be FN suggested that low coverage of the target space could weaken the results. Modification of the applied logical models should therefore be considered. Inclusion of knowledge of a higher percentage of the additional targets would lead to an improved reflection of reality and provide a more reliable basis to conclude if off-target effects enhance the prediction performance.

The final aim was to utilise the DrugProfiler tool to characterise the target profile of the drug *Imatinib* to substantiate the drug for off-label use in a cancer patient enrolled in the IMPRESS-Norway trial [68]. Analysis of the patient's molecular profile in comparison to the *Imatinib's* target profile implied that it could be an effective treatment option. The analysis demonstrates an

additional use case for the implemented module.

4.1 The Value of the Research

To properly understand how a drug effect is brought about, it is essential to posses knowledge of the molecules modified by the agent. The implemented module, accounting for off-target effects, enables exhaustive exploration of a drug's target space. Extensive drug target profiling is valuable when evaluating if a drug requires modifications to improve the selectively towards the target molecule of interest. The analysis can also show whether an agent is a candidate for repurposing. Furthermore, drug target profiling allows for enhanced *in silico* prediction of drug effects, providing a more accurate reflection of reality. Lastly, the knowledge is useful in precision medicine to identify if a drug has the potential to modify the gene products or signalling pathways found to be altered in a patient.

4.2 Limitations

Although valuable resources are implemented, and interesting results are obtained in this project, some limitations should be acknowledged. First, drugs are often screened towards a restricted number of possible targets, and as a result, the entire target space is not accounted for in the databases. Furthermore, binding affinity measurements are collected from multiple resources. Therefore, it is natural to expect variability in target predictions, which could lead to a weaker agreement between values measured between the same drug-target interaction (DTI). Finally, many of the identified drug targets were not included in the topology of the applied logical models, which hinders the accurate simulation of the drug effect.

4.3 Future Actions

Future actions could be pursued in regards to both the module and to improve the *in silico* simulation of drug effects. The drugs included in the database are now restricted to agents annotated for cancer. To increase the module's relevance across multiple diseases, drugs annotated for various conditions could be included. Furthermore, new entity and relationship types can be included to extend the database's application area.

The tool can also be developed to support additional queries. For instance, functionality can be implemented to retrieve drugs that modify one or multiple targets of interest. The identified drugs can be prioritised according to the number of targets of interest which they interact with and how strongly they bind to each target. The functionality will aid selection of drug treatment depending on a patients molecular profile. Additionally, it can facilitate drug repurposing and suggest off-label use.

For simulation of drug effects and prediction of drug synergies, it would be beneficial to modify the applied models to account for the identified off-targets bound with high affinity and with important functionality in the biological system of study. Improved models will likely give improved insight to how drug effects arise. Furthermore, it will enable improved reflection of reality, and as a result, raise the prediction performance of the DrugLogic pipeline. It would also be valuable to conduct analyses on additional logical models and using distinct drugs to confirm improved performance when off-targets are accounted for.

4.4 Concluding Summary

This study verifies the hypothesis that accounting for all known off-targeted effects of targeted therapies influences *in silico* drug response simulation. The findings emphasise the requirement for proper drug target profiling, accounting for off-targets. Such analysis is enabled through implementation of a module consisting of a database storing data on drug, targets and the prior knowledge available on drug-target interaction (DTI), and an associated tool facilitating automated drug target profiling. The use of target profiles, generated with the new module, is shown to provide improved insight of the drug effect through *in silico* simulation and enhanced prediction performance of drug synergies with the DrugLogic pipeline.

References

1. World Health Organization. WHO Report on Cancer: Setting Priorities, Investing Wisely and Providing Care for All. Accessed: April 18, 2023. Geneva: World Health Organization, 2020. Available from: https://www.who.int/publications/i/item/9789240001299
2. Wild CP, Weiderpass E and Stewart BW, eds. World Cancer Report: Cancer Research for Cancer Prevention. Accessed: April 18, 2023. Lyon: International Agency for Research on Cancer, 2020 Feb. Available from: http://publications.iarc.fr/586
3. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A and Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. eng. CA: a cancer journal for clinicians 2021; 71:209–49
4. Tran KB et al. The global burden of cancer attributable to risk factors, 2010–19: a systematic analysis for the Global Burden of Disease Study 2019. eng. The Lancet (British edition) 2022; 400:563–91
5. Krzyszczyk P, Acevedo A, Davidoff EJ, Timmins LM, Marrero-Berrios I, Patel M, White C, Lowe C, Sherba JJ, Hartmanshenn C et al. The growing role of precision and personalized medicine for cancer treatment. eng. Technology 2018; 6:79–100
6. Hanahan D and Weinberg RA. The Hallmarks of Cancer. eng. Cell 2000; 100:57–70
7. Carlberg C. Cancer biology : how science works. eng. Cham, Switzerland: Springer, 2021
8. Hanahan D and Weinberg RA. Hallmarks of Cancer: The Next Generation. eng. Cell 2011; 144:646–74
9. Hanahan D. Hallmarks of Cancer: New Dimensions. eng. Cancer discovery 2022; 12:31–46
10. Bruce Alberts John H. Wilson TH. Molecular biology of the cell. eng. 6th ed. New York: Garland Science, 2015
11. Dancey JE and Chen HX. Strategies for optimizing combinations of molecularly targeted anticancer agents. eng. Nature reviews. Drug discovery 2006; 5:649–59
12. Hou J, He Z, Liu T, Chen D, Wang B, Wen Q and Zheng X. Evolution of Molecular Targeted Cancer Therapy: Mechanisms of Drug Resistance and Novel Opportunities Identified by CRISPR-Cas9 Screening. eng. Frontiers in oncology 2022; 12:755053–3
13. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. eng. Nature chemical biology 2008; 4:682–90
14. Kaufmann SHE. Paul Ehrlich: founder of chemotherapy. eng. Nature reviews. Drug discovery 2008; 7:373–3
15. Lee YT, Tan YJ and Oon CE. Molecular targeted therapy: Treating cancer with specificity. eng. European journal of pharmacology 2018; 834:188–96
16. Weiner LM, Surana R and Wang S. Monoclonal antibodies: versatile platforms for cancer immunotherapy. eng. Nature reviews. Immunology 2010; 10:317–27
17. Padma VV. An overview of targeted cancer therapy. eng. Biomedicine (Taipei) 2015; 5:1–6
18. Druker BJ, Guilhot F, O'Brien SG, Gathmann I, Kantarjian H, Gattermann N, Deininger MW, Silver RT, Goldman JM, Stone RM, Cervantes F, Hochhaus A, Powell BL, Gabrilove JL, Rousselot P, Reiffers J, Cornelissen JJ, Hughes T, Agis H, Fischer T, Verhoef G, Shepherd J, Saglio G, Gratwohl A, Nielsen JL, Radich JP, Simonsson B, Taylor K, Baccarani M, So C, Letvak L and Larson RA. Five-Year Follow-up of Patients Receiving Imatinib for Chronic Myeloid Leukemia. eng. The New England journal of medicine 2006; 355:2408–17
19. Housman G, Byler S, Heerboth S, Lapinska K, Longacre M, Snyder N and Sarkar S. Drug resistance in cancer: An overview. eng. Cancers 2014; 6:1769–92

20. Rudmann DG. On-target and Off-target-based Toxicologic Effects. eng. Toxicologic pathology 2013; 41:310–4

21. Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, Hocker M, Treiber DK and Zarrinkar PP. Comprehensive analysis of kinase inhibitor selectivity. eng. Nature biotechnology 2011; 29:1046–51

22. Morphy R. Selectively Nonselective Kinase Inhibition: Striking the Right Balance. eng. Journal of medicinal chemistry 2010; 53:1413–37

23. Knight JDR, Qian B, Baker D and Kothary R. Conservation, variability and the modeling of active protein kinases. eng. PloS one 2007; 2:e982–e982

24. Verkhivker GM. Exploring sequence-structure relationships in the tyrosine kinome space: functional classification of the binding specificity mechanisms for cancer therapeutics. eng. Bioinformatics 2007; 23:1919–26

25. Force T and Kolaja KL. Cardiotoxicity of kinase inhibitors: the prediction and translation of preclinical models to clinical outcomes. eng. Nature reviews. Drug discovery 2011; 10:111–26

26. Morphy R, Kay C and Rankovic Z. From magic bullets to designed multiple ligands. eng. Drug Discovery Today 2004; 9:641–51

27. Wermuth CG. Multitargeted drugs: the end of the 'one-target-one-disease' philosophy? eng. Drug discovery today 2004; 9:826–7

28. Kitano H. Towards a theory of biological robustness. eng. Molecular systems biology 2007; 3:137

29. Flobak Å, Baudot A, Remy E, Thommesen L, Thieffry D, Kuiper M and Lægreid A. Discovery of Drug Synergies in Gastric Cancer Cells Predicted by Logical Modeling. eng. PLoS computational biology 2015; 11:e1004426–e1004426

30. Niederdorfer B, Touré V, Vazquez M, Thommesen L, Kuiper M, Lægreid A and Flobak Å. Strategies to Enhance Logic Modeling-Based Cell Line-Specific Drug Synergy Prediction. eng. Frontiers in physiology 2020; 11:862–2

31. Larkin J, Ascierto PA, Dréno B, Atkinson V, Liszkay G, Maio M, Mandalà M, Demidov L, Stroyakovskiy D, Thomas L, Cruz-Merino L de la, Dutriaux C, Garbe C, Sovak MA, Chang I, Choong N, Hack SP, McArthur GA and Ribas A. Combined Vemurafenib and Cobimetinib in BRAF-Mutated Melanoma. eng. The New England journal of medicine 2014; 371:1867–76

32. Cristofanilli M, Turner NC and Bondarenko. Fulvestrant plus palbociclib versus fulvestrant plus placebo for treatment of hormone-receptor-positive, HER2-negative metastatic breast cancer that progressed on previous endocrine therapy (PALOMA-3): final analysis of the multicentre, double-blind, phase 3 randomised controlled trial (vol 17, pg 431, 2016). eng. The lancet oncology 2016; 17:E270–E270

33. Pahikkala T, Airola A, Pietilä S, Shakyawar S, Szwajda A, Tang J and Aittokallio T. Toward more realistic drug–target interaction predictions. eng. Briefings in Bioinformatics 2014 Apr; 16:325–37

34. Nelson DL. Lehninger principles of biochemistry. eng. 6th ed. New York: W.H. Freeman, 2013

35. Shim J, Hong ZY, Sohn I and Hwang C. Prediction of drug-target binding affinity using similarity-based convolutional neural network. eng. Scientific reports 2021; 11:4416–6

36. Öztürk H, Özgür A and Ozkirimli E. DeepDTA: deep drug-target binding affinity prediction. eng. Bioinformatics 2018; 34:i821–i829

37. Cer RZ, Mudunuri U, Stephens R and Lebeda FJ. IC50-to-Ki: a web-based tool for converting IC50 to Ki values for inhibitors of enzyme activity and ligand binding. eng. Nucleic acids research 2009; 37:W441–W445

38. Trewavas A. A brief history of systems biology - "Every object that biology studies is a system of systems." Francois Jacob (1974). eng. The Plant cell 2006; 18:2420–30

39. Kitano H. Systems Biology: A Brief Overview. eng. Science (American Association for the Advancement of Science) 2002; 295:1662–4

40. Westerhoff H and Palsson B. The evolution of molecular biology into systems biology. eng. Nature biotechnology 2004; 22:1249–52

41. Auffray C, Chen Z and Hood L. Systems medicine: The future of medical genomics and healthcare. eng. Genome medicine 2009; 1:2

42. Hornberg JJ, Bruggeman FJ, Westerhoff HV and Lankelma J. Cancer: A Systems Biology disease. eng. BioSystems 2006; 83:81–90

43. Azeloglu EU and Iyengar R. Signaling networks: Information flow, computation, and decision making. eng. Cold Spring Harbor perspectives in biology 2015; 7:a005934–a005934

44. Barabási AL and Oltvai ZN. Network biology: understanding the cell's functional organization. eng. Nature reviews. Genetics 2004; 5:101–13

45. Albert I, Thakar J, Li S, Zhang R and Albert R. Boolean network simulations for life scientists. eng. Source code for biology and medicine 2008; 3:16–6

46. Klipp E, Liebermeister W, Wierling C, Kowald A, Lehrach H and Herwig R. Systems biology : a textbook. eng. 1st ed. Weinheim: Wiley, 2009

47. Naldi A, Remy E, Thieffry D and Chaouiya C. Dynamically consistent reduction of logical regulatory graphs. eng. Theoretical computer science 2011; 412:2207–18

48. Abou-Jaoudé W, Traynard P, Monteiro PT, Saez-Rodriguez J, Helikar T, Thieffry D and Chaouiya C. Logical modeling and dynamical analysis of cellular networks. eng. Frontiers in genetics 2016; 7:94–4

49. Tsirvouli E, Touré V, Niederdorfer B, Vázquez M, Flobak Å and Kuiper M. A Middle-Out Modeling Strategy to Extend a Colon Cancer Logical Model Improves Drug Synergy Predictions in Epithelial-Derived Cancer Cell Lines. eng. Frontiers in molecular biosciences 2020; 7:502573–3

50. Elmasri R. Fundamentals of database systems. eng. 7th ed. Boston , Mass: Pearson, 2016

51. Consortium TU. UniProt: the universal protein knowledgebase. eng. Nucleic Acids Research 2016; 45:D158–D169

52. Organisation HHG. Human Genome Organisation. [Internet]. [updated 2023 May 6; cited 2023 May 10]. Cambridge, UK, 2023 May. Available from: https://www.hugo-international.org/about-us/

53. Ochoa D, Hercules A, Carmona M, Suveges D, Baker J, Malangone C, Lopez I, Miranda A, Cruz-Castillo C, Fumis L, Bernal-Llinares M, Tsukanov K, Cornu H, Tsirigos K, Razuvayevskaya O, Buniello A, Schwartzentruber J, Karim M, Ariano B, Martinez Osorio RE, Ferrer J, Ge X, Machlitt-Northen S, Gonzalez-Uriarte A, Saha S, Tirunagari S, Mehta C, Roldán-Romero JM, Horswell S, Young S, Ghoussaini M, Hulcoop DG, Dunham I and McDonagh EM. The next-generation Open Targets Platform: reimagined, redesigned, rebuilt. eng. Nucleic acids research 2023; 51:D1353–D1359

54. Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, Magariños MP, Mosquera JF, Mutowo P, Nowotka M, Gordillo-Marañón M, Hunter F, Junco L, Mugumbate G, Rodriguez-Lopez M, Atkinson F, Bosc N, Radoux CJ, Segura-Cabrera A, Hersey A and Leach AR. ChEMBL: towards direct deposition of bioassay data. eng. Nucleic acids research 2019; 47:D930–D940

55. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, Bindal N, Beare D, Smith JA, Thompson IR, Ramaswamy S, Futreal PA, Haber DA, Stratton MR, Benes C, McDermott U and Garnett MJ. Genomics of Drug Sensitivity in Cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. eng. Nucleic acids research 2013; 41:D955–D961

56. Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L and Chong J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. eng. Nucleic acids research 2016; 44:D1045–D1053

57. Tang J, Karhinen L, Xu T, Szwajda A, Yadav B, Wennerberg K and Aittokallio T. Target Inhibition Networks: Predicting Selective Combinations of Druggable Targets to Block Cancer Survival Pathways. eng. PLoS computational biology 2013; 9:e1003226

58. Flobak Å, Niederdorfer B, Nakstad VT, Thommesen L, Klinkenberg G and Lægreid A. A high-throughput drug combination screen of targeted small molecule inhibitors in cancer cell lines. eng. Scientific data 2019; 6:237–10

59. Tang J, Szwajda A, Shakyawar S, Xu T, Hintsanen P, Wennerberg K and Aittokallio T. Making Sense of Large-Scale Kinase Inhibitor Bioactivity Data Sets: A Comparative and Integrative Analysis. eng. Journal of Chemical Information and Modeling 2014; 54:735–43

60. Eid S, Turk S, Volkamer A, Rippmann F and Fulle S. Kinmap: A web-based tool for interactive navigation through human kinome data. eng. BMC bioinformatics 2017; 18:16–6

61. Perfetto L, Briganti L, Calderone A, Perpetuini AC, Iannuccelli M, Langone F, Licata L, Marinkovic M, Mattioni A, Pavlidou T, Peluso D, Petrilli LL, Pirró S, Posca D, Santonico E, Silvestri A, Spada F, Castagnoli L and Cesareni G. SIGNOR: A database of causal relationships between biological entities. eng. Nucleic acids research 2016; 44:D548–D554

62. Kanehisa M and Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. eng. Nucleic acids research 2000; 28:27–30

63. Stoll G, Caron B, Viara E, Dugourd A, Zinovyev A, Naldi A, Kroemer G, Barillot E and Calzone L. MaBoSS 2.0: An environment for stochastic Boolean modeling. eng. BIOINFORMATICS 2017; 33:2226–8

64. Naldi A, Hernandez C, Levy N, Stoll G, Monteiro PT, Chaouiya C, Helikar T, Zinovyev A, Calzone L, Cohen-Boulakia S, Thieffry D and Paulevé L. The CoLoMoTo interactive notebook: Accessible and reproducible computational analyses for qualitative biological networks. eng. Frontiers in physiology 2018; 9:680–0

65. Levy N, Naldi A, Hernandez C, Stoll G, Thieffry D, Zinovyev A, Calzone L and Paulevé L. Prediction of mutations to control pathways enabling tumor cell invasion with the CoLoMoTo Interactive Notebook (Tutorial). eng. Frontiers in physiology 2018; 9:787–7

66. Berenbaum MC. What is synergy? eng. Pharmacological Reviews 1989; 41:93–141

67. Zobolas J, Kuiper M and Flobak Å. emba: R package for analysis and visualization of biomarkers in boolean model ensembles. eng. Journal of open source software 2020; 5:2583

68. Helland Å, Russnes HG, Fagereng GL, Al-Shibli K, Andersson Y, Berg T, Bjørge L, Blix E, Bjerkehagen B, Brabrand S, Cameron MG, Dalhaug A, Dietzel D, Dønnem T, Enerly E, Flobak Å, Fluge S, Gilje B, Gjertsen BT, Grønberg BH, Grønås K, Guren T, Hamre H, Haug Å, Heinrich D, Hjortland GO, Hovig E, Hovland R, Iversen AC, Janssen E, Kyte JA, Lippe Gythfeldt H von der, Lothe R, Lund JÅ, Meza-Zepeda L, Munthe-Kaas MC, Nguyen OTD, Niehusmann P, NilsenPuco HK, Ree AH, Riste TB, Semb K, Steinskog ESS, Stensvold A, Suhrke P, Tennøe Ø, Tjønnfjord GE, Vassbotn LJ, Aas E, Aasebø K, Tasken K

and Smeland S. Improving public cancer care by implementing precision medicine in Norway: IMPRESS-Norway. eng. Journal of translational medicine 2022; 20:225–5

69. Gilson MK. An Introduction to Protein-Ligand Binding for BindingDB Users. [Internet]. [updated 2010 December 15; cited 2023 May 1]. San Diego, US, 2010 Dec. Available from: https://www.bindingdb.org/bind/BindingDB-Intro2a.pdf

70. Keenan AB et al. The Library of Integrated Network-Based Cellular Signatures NIH Program: System-Level Cataloging of Human Cells Response to Perturbations. eng. Cell systems 2018; 6:13–24

71. Yang M, Jaaks P, Dry J, Garnett M, Menden MP and Saez-Rodriguez J. Stratification and prediction of drug synergy based on target functional similarity. eng. NPJ systems biology and applications 2020; 6:16–6

72. Flobak Å, Zobolas J, Vazquez M, Steigedal TS, Thommesen L, Grislingås A, Niederdorfer B, Folkesson E and Kuiper M. Logical modeling: Combining manual curation and automated parameterization to predict drug synergies. eng. bioRxiv 2021

73. Zarrinkar PP, Karaman MW, Herrgard S, Treiber DK, Gallant P, Atteridge CE, Campbell BT, Chan KW, Ciceri P, Davis MI, Edeen PT, Faraoni R, Floyd M, Hunt JP, Lockhart DJ, Milanov ZV, Morrison MJ, Pallares G, Patel HK, Pritchard S and Wodicka LM. A quantitative analysis of kinase inhibitor selectivity. eng. Nature biotechnology 2008; 26:127–32

74. Samsdodd F. Target-based drug discovery: is something wrong? eng. Drug Discovery Today 2005; 10:139–47

75. Klaeger S, Heinzlmeir S, Wilhelm M, Polzer H, Vick B, Koenig PA, Reinecke M, Ruprecht B, Petzoldt S, Meng C, Zecha J, Reiter K, Qiao H, Helm D, Koch H, Schoof M, Canevari G, Casale E, Re Depaolini S, Feuchtinger A, Wu Z, Schmidt T, Rueckert L, Becker W, Huenges J, Garz AK, Gohlke BO, Zolg DP, Kayser G, Vooder T, Preissner R, Hahne H, Tõnisson N, Kramer K, Götze K, Bassermann F, Schlegl J, Ehrlich HC, Aiche S, Walch A, Greif PA, Schneider S, Felder ER, Ruland J, Médard G, Jeremias I, Spiekermann K and Kuster B. The target landscape of clinical kinase drugs. eng. Science (American Association for the Advancement of Science) 2017; 358:1148–8

76. Morin M. From oncogene to drug: Development of small molecule tyrosine kinase inhibitors as anti-tumor and anti-angiogenic agents. eng. Oncogene 2000; 19:6574–83

77. Zhao Y, Zhang X, Chen Y, Lu S, Peng Y, Wang X, Guo C, Zhou A, Zhang J, Luo Y, Shen Q, Ding J, Meng L and Zhang J. Crystal Structures of PI3Kaplha Complexed with PI103 and Its Derivatives: New Directions for Inhibitors Design. eng. ACS medicinal chemistry letters 2014; 5:138–42

78. Williams R, Berndt A, Miller S, Hon WC and Zhang X. Form and flexibility in phosphoinositide 3-kinases. eng. Biochemical Society Transactions 2009 Jul; 37:615–26

79. Sapkota GP, Cummings L, Newell FS, Armstrong C, Bain J, Frodin M, Grauert M, Hoffmann M, Schnapp G, Steegmaier M, Cohen P and Alessi DR. BI-D1870 is a specific inhibitor of the p90 RSK (ribosomal S6 kinase) isoforms in vitro and in vivo. eng. Biochemical journal 2007; 401:29–38

80. Kesarwani M, Huber E, Kincaid Z, Evelyn CR, Biesiada J, Rance M, Thapa MB, Shah NP, Meller J, Zheng Y and Azam M. Targeting substrate-site in Jak2 kinase prevents emergence of genetic resistance. eng. Scientific reports 2015; 5:14538–8

81. Kuma Y, Sabio G, Bain J, Shpiro N, Márquez R and Cuenda A. BIRB796 Inhibits All p38 MAPK Isoforms in Vitro and in Vivo. eng. The Journal of biological chemistry 2005; 280:19472–9

82. Pargellis C, Tong L, Churchill L, Cirillo PF, Gilmore T, Graham AG, Grob PM, Hickey ER, Moss N, Pav S and Regan J. Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. eng. Nature structural biology 2002; 9:268–72

83. Nagar B. C-abl tyrosine kinase and inhibition by the cancer drug imatinib (Gleevec/STI-571). eng. The Journal of nutrition 2007; 137:1518S–1523S

84. Ellestad GA. (5Z)-7-Oxozeaenol: A novel and potent resorcylic acid lactone kinase inhibitor with a cis-enone Michael acceptor. eng. Chirality (New York, N.Y.) 2019; 31:110–7

85. Winssinger N and Barluenga S. Chemistry and biology of resorcylic acid lactones. eng. Chem. Commun. 2007; 1:22–36

86. Breen ME and Soellner MB. Small Molecule Substrate Phosphorylation Site Inhibitors of Protein Kinases: Approaches and Challenges. eng. ACS chemical biology 2015; 10:175–89

87. Barnett SF, Defeo-Jones D, Fu S, Hancock PJ, Haskell KM, Jones RE, Kahana JA, Kral AM, Leander K, Lee LL, Malinowski J, McAvoy EM, Nahas DD, Robinson RG and Huber HE. Identification and characterization of pleckstrin-homology-domain-dependent and isoenzyme-specific Akt inhibitors. eng. Biochemical journal 2005; 385:399–408

88. Anastassiadis T, Deacon SW, Devarajan K, Ma H and Peterson JR. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. eng. Nature biotechnology 2011; 29:1039–45

89. Cohen P. Protein kinases - the major drug targets of the twenty-first century? eng. Nature reviews. Drug discovery 2002; 1:309–15

90. Duran CL, Borriello L, Karagiannis GS, Entenberg D, Oktay MH and Condeelis JS. Targeting tie2 in the tumor microenvironment: From angiogenesis to dissemination. eng. Cancers 2021; 13:5730

91. Jones N and Dumont DJ. Tek/Tie2 signaling : New and old partners: Tumor Angiogenesis and Anti-Angiogenic Therapy. eng. Cancer and metastasis reviews 2000; 19:13–7

92. Kohl P, Crampin EJ, Quinn TA and Noble D. Systems Biology: An Approach. eng. Clinical pharmacology and therapeutics 2010; 88:25–33

93. McNew KL, Whipple WJ, Mehta AK, Grant TJ, Ray L, Kenny C and Singh A. MEK and TAK1 regulate apoptosis in colon cancer cells with KRAS-dependent activation of proinflammatory signaling. eng. Molecular cancer research 2016; 14:1204–16

94. Kim IJ, Park JH, Kang HC, Shin Y, Park HW, Park HR, Ku JL, Lim SB and Park JG. Mutational analysis of BRAF and K-ras in gastric cancers: Absence of BRAF mutations in gastric cancers. eng. Human genetics 2003; 114:118–20

95. Zambrowicz BP and Sands AT. Modeling drug action in the mouse with knockouts and RNA interference. eng. Drug Discovery Today: Targets 2004; 3:198–207

96. Winzeler E, Shoemaker D, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke J, Bussey H, Chu A and Connelly C. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. eng. Science (American Association for the Advancement of Science) 1999; 285:901–6

97. Davis RW, Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, Dow S, Lucau-Danila A, Anderson K, André B, Arkin AP, Astromoff A, El Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, Entian KD, Flaherty P, Foury F, Garfinkel DJ, Gerstein M, Gotte D, Güldener U, Hegemann JH, Hempel S, Herman Z, Jaramillo DF, Kelly DE, Kelly SL, Kötter P, LaBonte D, Lamb DC, Lan N, Liang H, Liao H, Liu L, Luo C, Lussier M, Mao R, Menard P, Ooi SL, Revuelta JL, Roberts CJ, Rose M, Ross-Macdonald P, Scherens B, Schimmack G, Shafer B, Shoemaker DD,

Sookhai-Mahadeo S, Storms RK, Strathern JN, Valle G, Voet M, Volckaert G, Wang Cy, Ward TR, Wilhelmy J, Winzeler EA, Yang Y, Yen G, Youngman E, Yu K, Bussey H, Boeke JD, Snyder M, Philippsen P and Johnston M. Functional profiling of the Saccharomyces cerevisiae genome. eng. Nature (London) 2002; 418:387–91

98. Groom CR and Hopkins AL. The druggable genome. eng. Nature reviews. Drug discovery 2002; 1:727–30

99. Petrelli A and Giordano S. From Single- to Multi-Target Drugs in Cancer Therapy: When Aspecificity Becomes an Advantage. eng. Current medicinal chemistry 2008; 15:422–32

100. Dessalew N and Mikre W. On the paradigm shift towards multitarget selective drug design. eng. Current Computer-Aided Drug Design 2008; 4:76–90

101. Barabási AL, Jeong H and Albert R. Error and attack tolerance of complex networks. eng. Nature (London) 2000; 406:378–82

102. Iqbal N and Iqbal N. Imatinib: A Breakthrough of Targeted Therapy in Cancer. eng. Chemotherapy research and practice 2014; 2014:357027–9

103. Demetri GD, Mehren M von, Blanke CD, Van den Abbeele AD, Eisenberg B, Roberts PJ, Heinrich MC, Tuveson DA, Singer S, Janicek M, Fletcher CD, Fletcher JA, Silverman SG, Silberman SL, Capdeville R, Kiese B, Peng B, Dimitrijevic S, Druker BJ, Corless C and Joensuu H. Efficacy and Safety of Imatinib Mesylate in Advanced Gastrointestinal Stromal Tumors. eng. The New England journal of medicine 2002; 347:472–80

104. Carvajal RD, Antonescu CR, Wolchok JD, Chapman PB, Roman RA, Teitcher J, Panageas KS, Busam KJ, Chmielowski B, Lutzky J, Pavlick AC, Fusco A, Cane L, Takebe N, Vemula S, Bouvier N, Bastian BC and Schwartz GK. KIT as a Therapeutic Target in Metastatic Melanoma. eng. JAMA : the journal of the American Medical Association 2011; 305:2327–34

Appendix A.1 - Tools and Data Resources

To enable reproducibility of the work conducted in this project, the versions of the applied programs and tools are listed in Table A.1. Additionally, the versions of the Python and R packages used in the programs are listed in Table A.2 and Table A.3, respectively. Finally, Table A.4 provides the date for the latest download of the used data resources.

Table A.1: Applied version of programs and tools used in current project.

Programs and Tools	**Version**
Python	3.8.5
R	1.3.1073
DB Browser for SQLite	3.12.2
Docker	4.4.2 (73305)
CoLoMoTo Interactive Notebook	2022-07-01
DrugLogics software	1.2.1 (Docker image 1.0)
gitsbe	1.3.1
drabme	1.2.1

Table A.2: Applied version of Python packages used in current project.

Library	**Version**
pandas	1.1.3
sqlite3	2.6.0
dash	2.7.0
dash-cytoscape	0.2.0
maboss	0.7.19
matplotlib	3.3.2
scipy	1.5.2
maboss	?

Table A.3: Applied version of R packages used in current project.

Library	**Version**
dplyr	1.1.0
tibble	3.1.7
emba	0.1.8
usefun	0.4.8
PRROC	1.3.1
DT	0.23

Table A.4: Date for latest download of data from applied resources.

Data resource	**Version**
UniProtKB	24.11.22
ChEMBL	03.02.23
OpenTargets	21.09.22
BindingDB	19.10.22
Genomics of Drug Sensitivity in Cancer	12.10.22

Appendix A.2 - Chemical Inhibitores

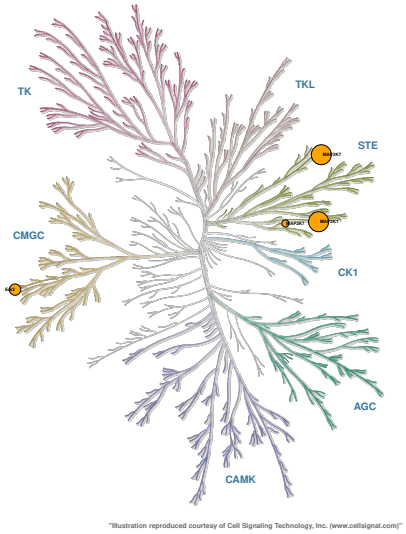
The chemical inhibitors applied in the study are listed in Table A.5 along with the unique identifier, primary targets, formula and InChiKey.

Table A.5: Inhibitor name, ChEMBL ID, primary target, formula and InChiKey for the chemical inhibitors.

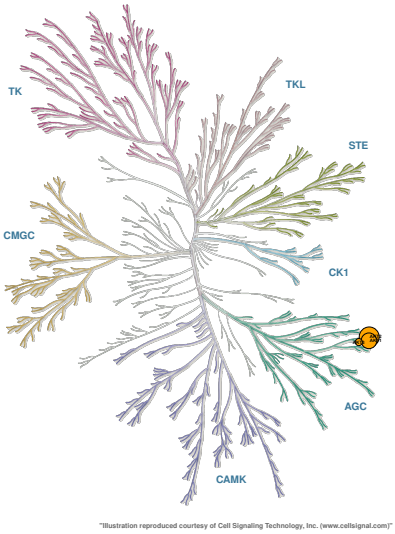
Chemical inhibitor	ChEMBL ID	Primary target	Formula	InChiKey
(5Z)-7-oxozeaenol (LL-Z1640-2)	CHEMBL1077979	MAP3K7	C19H22O7	NEQZWEXWOFPKOT-BYRRXHGESA-N
AKTi-1,2 (AKT inhibitor VIII)	CHEMBL258844	AKT1, AKT2, AKT3	C34H29N7O	IWCQHVUQEFDRIW-UHFFFAOYSA-N
BIRB0796 (Dormapimod)	CHEMBL103667	MAPK14	C31H37N5O3	MVCOAUNKQVWQHZ-UHFFFAOYSA-N
CT99021 (CHIR 99021)	CHEMBL412142	GSK3A, GSK3B	C22H18Cl2N8	AQGNHMOJWBZFQQ-UHFFFAOYSA-N
PD0325901	CHEMBL507361	MAP2K1, MAP2K2	C16H14F3IN2O4	SUDAHWBOROXANE-SECBINFHSA-N
PI103	CHEMBL573339	PIK3CA, PIK3CB	C19H16N4O3	TUVCWJQQGGETHL-UHFFFAOYSA-N
PKF118-310 (Toxoflavin)	CHEMBL578512	CTNNB1	C7H7N5O2	SLGRAIAQIAUZAQ-UHFFFAOYSA-N
JNK-IN-8 (JNK Inhibitor XVI)	CHEMBL2216824	MAPK8, MAPK9, MAPK10	C29H29N7O2	GJFCSAPFHAXMSF-UXBLZVDNSA-N
BI-D1870	CHEMBL573107	RPS6KA1, RPS6KA3, RPS6KA2, RPS6KA6	C19H23F2N5O2	DTEKTGDVSARYDS-UHFFFAOYSA-N
BI605906 (BIX02514)	CHEMBL4522930	IKBKB	C17H22F2N4O3S2	IYHHRZBKXXKDDY-UHFFFAOYSA-N
Ruxolitinib (INCB18424)	CHEMBL1789941	JAK1, JAK2	C17H18N6	HFNKQEVNSGCOJV-OAHLLOKOSA-N
SB-505124	CHEMBL1824446	TGFBR1, ACVR1B, ACVR1C	C20H21N3O2 · xHCl · yH2O	DIDCCMVWCVRTNB-UHFFFAOYSA-N
D4476	CHEMBL410456	CSNK1D, TGFBR1	C23H18N4O3	DPDZHVCKYBCJHW-UHFFFAOYSA-N
10058-F4	CHEMBL1568415	MYC	C12H11NOS2	SVXDHPADAXBMFB-JXMROGBWSA-N
Stattic	CHEMBL1337170	STAT3	C8H5NO4S	ZRRGOUHITGRLBA-UHFFFAOYSA-N
GSK2334470	CHEMBL1765740	PDPK1	C25H34N8O	QLPHOXTXAKOFMU-WBVHZDCISA-N
PRT062607 (P505-15)	CHEMBL2177736	SYK	C19H23N9O	TXGKRVFSSHPBAJ-JKSUJKDBSA-N

Appendix A.3 - Kinase Maps

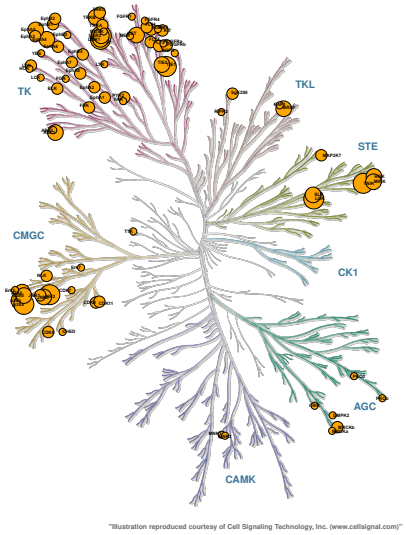
Phylogenetic trees of the human kinome were generated for the kinase inhibitors of study and annotated with their target profile. The annotation size was scaled according to the lowest binding affinity value measured between the drug-target interaction. The result is presented in Figure A.1.



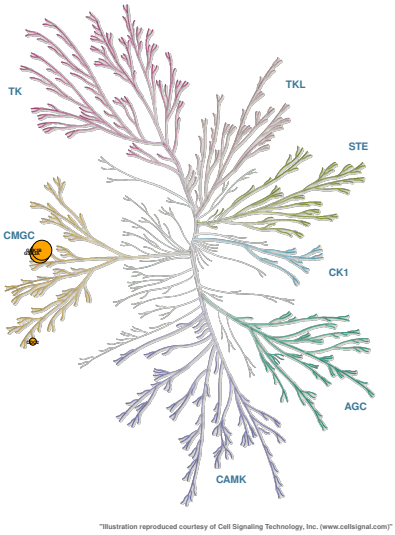
(a) (5Z)-7-oxozeaneol



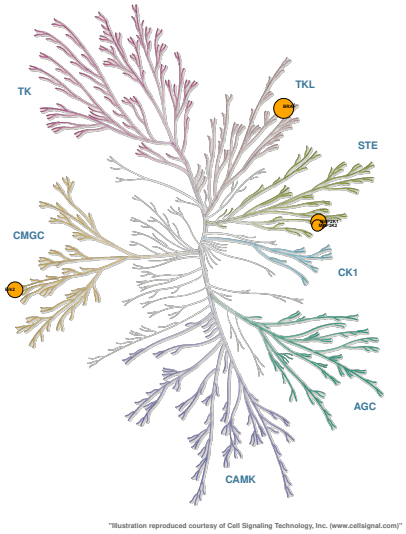
(b) AKTi1,2



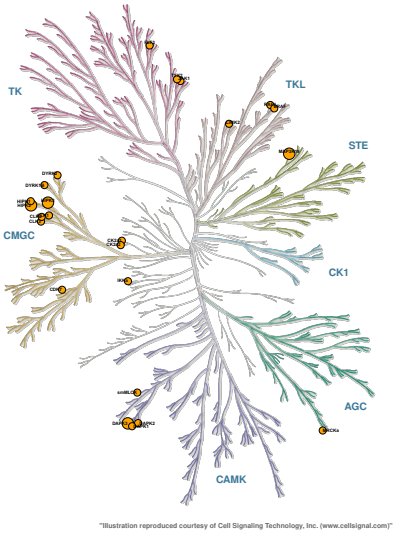
(c) BIRB0796



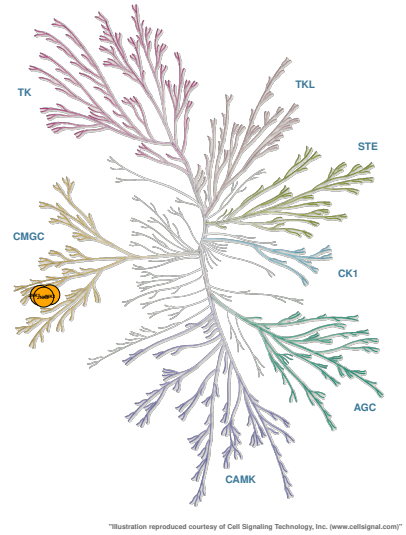
(d) CT99021



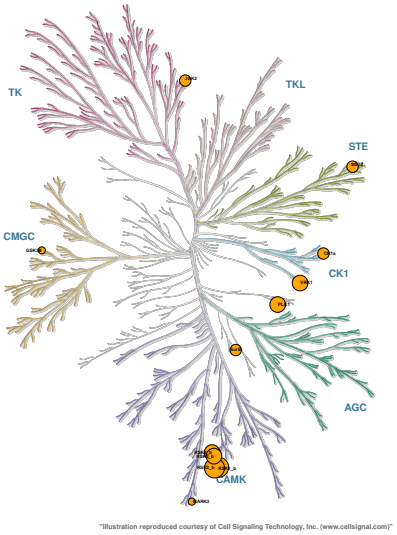
(e) PD0325901



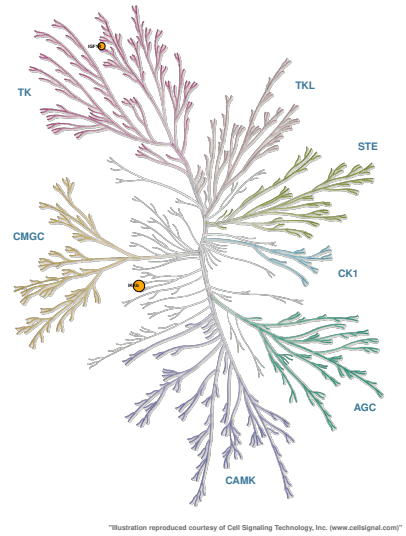
(f) PI103



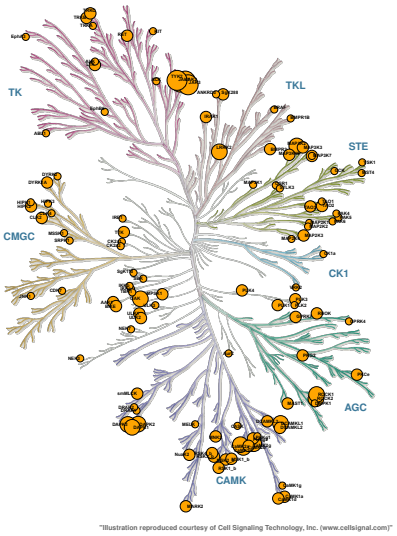
(g) JNK-IN-8



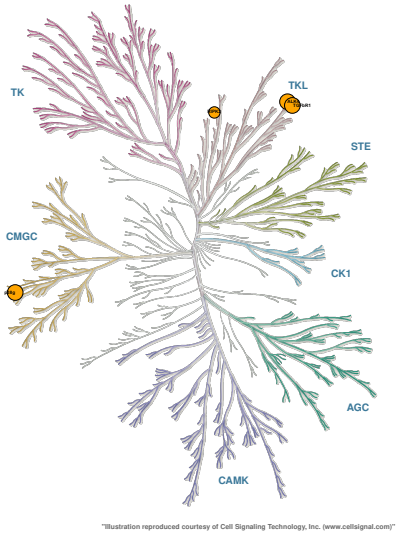
(h) BI-D1870



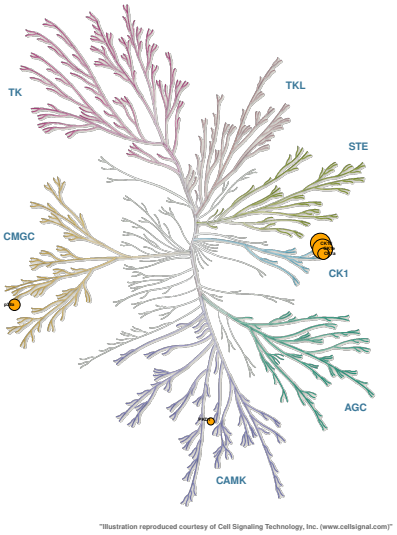
(i) BI605906



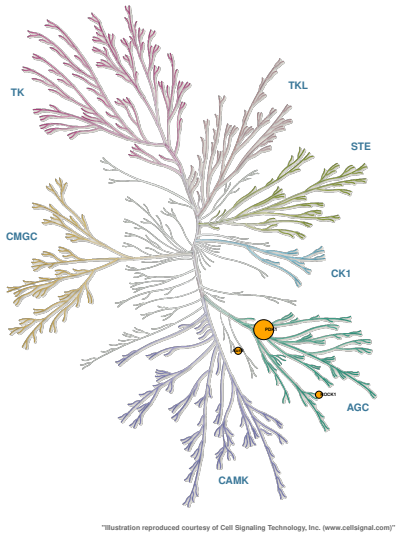
(j) Ruxolitinib



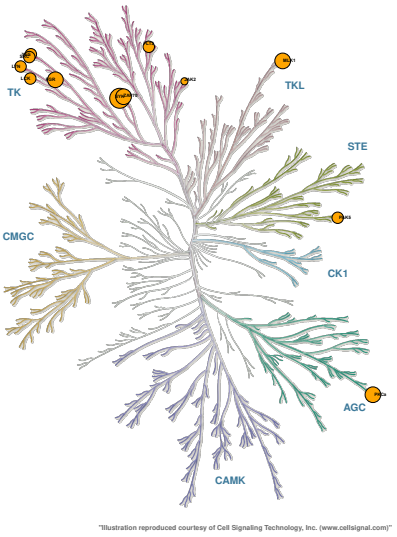
(k) SB-505124



(l) D4476



(m) GSK2334470



(n) PRT062607

Figure A.1: Phylogenetic trees of the human kinome generated by KinMap, and annotated with inhibition profiles of the kinase inhibitors a) (5Z)-7-oxozeaneol, b) AKTi1,2, c) BIRB0796, d) CT99021, e) PD0325901, f) PI103, g) JNK-IN-8, h) BI-D1870, i) BI605906, j) Ruxolitinib, k) SB-505124, l) D4476, m) GSK2334470, and n) PRT062607. The circle size indicates the binding affinity strength of the drug target pair, measured <10 nM, 10 nM to 100 nM, 100 nM to 1 µM, or 1 µM to 10 µM.

Appendix A.4 - Filtrated Target Profiles

The DrugProfiler was applied to generate target profiles for 17 chemical inhibitors, using four different cut-offs for binding affinity. The target profiles were modified for simulations of CAncer Signaling CAusality DatabasE (CASCADE) 1.0 and CASCADE 2.0 by excluding the components not accounted for in these logical models. The name of the chemical inhibitors, their primary target and profiles updated according to CASCADE 1.0 and CASCADE 2.0 are presented in Tabel A.6 and Table A.7, respectively.

Table A.6: Name of chemical inhibitor, primary target and target profiles generated by the DrugProfiler tool. The applied thresholds for dissociation constant (K_d), inhibition constant (K_i), and half maximal inhibitory concentration (IC_{50}) were 10 nM, 100 nM, 1 µM, and 10 µM. The profiles are modified to only include proteins in the CASCADE 1.0 network model.

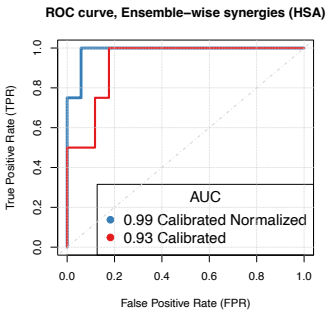
Chemical inhibitor	**Primary target**	**10 nM**	**100 nM**	**1 µM**	**10 µM**
(5Z)-7-oxozeaenol	MAP3K7	MAP2K1, MAP3K7	MAP2K1, MAP3K7	MAP2K1, MAP3K7, MAPK1	MAP2K1, MAP2K7, MAP3K7, MAPK1
AKTi-1,2	AKT1, AKT2	AKT2	AKT1, AKT2	AKT1, AKT2	AKT1, AKT2, AKT3
BIRB0796	MAPK14	MAPK14, MAPK9	BRAF, MAPK14, MAPK9	BRAF, FLT1 FLT3, FLT4 KDR, KIT, MAP3K7, MAPK14 MAPK8, MAPK9 NLK, PDGFRB	BRAF, CSF1R, EGFR, FGFR1, FGFR4, FLT1, FLT3, FLT4, KDR, KIT, MAP3K7, MAPK1, MAPK14, MAPK8, MAPK9, NLK, PDGFRA, PDGFRB, RAF1
CT99021	GSK3A, GSK3B	GSK3A, GSK3B	GSK3A, GSK3B	GSK3A, GSK3B	GSK3A, GSK3B
PD0325901	MAP2K1, MAP2K2	BRAF, MAP2K1	BRAF, MAP2K1, MAPK1	BRAF, MAP2K1, MAP2K2, MAPK1	BRAF, MAP2K1, MAP2K2, MAPK1
PI103	PIK3CA	PIK3CA, MTOR	PIK3CA, MTOR, MLST8	PIK3CA, MTOR, MLST8	BRAF, CHUK, FLT3, MLST8, MTOR, PIK3CA, RAF1
PKF118-310	CTNNB1	-	-	DUSP1	CTNNB1, DUSP1, TCF7L2

Table A.7: Name, primary target and target profiles for analysed chemical inhibitors. The target profiles were generated by the DrugProfiler tool. The applied thresholds for dissociation constant (K_d), inhibition constant (K_i), and half maximal inhibitory concentration (IC_{50}) were 10 nM, 100 nM, 1 µM, and 10 µM. Each target profile is modified to only consist of components included in the CASCADE 2.0 model

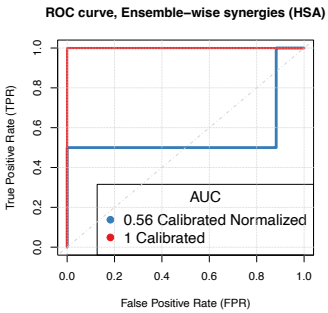
Chemical inhibitor	Primary target	10 nM	100 nM	1 µM	10 µM
(5Z)-7-oxozeaenol	MAP3K7	MAP2K1, MAP3K7	MAP2K1, MAP3K7	MAP2K1, MAP3K7, MAPK1	MAP2K1, MAP2K7, MAP3K7, MAPK1
AKTi-1,2	AKT1, AKT2	AKT2	AKT1, AKT2	AKT1, AKT2	AKT1, AKT2, AKT3
BIRB0796	MAPK14	MAPK14, MAPK9	BRAF, MAPK14, MAPK9	BRAF, FLT1, FLT3, FLT4, KDR, KIT, MAP3K7, MAPK14, MAPK8, MAPK9, NLK, PDGFRB	BRAF, CSF1R, EGFR, FGFR1, FGFR4, FLT1, FLT3, FLT4, KDR, KIT, MAP3K7, MAPK1, MAPK14, MAPK8, MAPK9, NLK, PDGFRA, PDGFRB, PRKCD, RAF1
CT99021	GSK3A, GSK3B	GSK3A, GSK3B	GSK3A, GSK3B	GSK3A, GSK3B	GSK3A, GSK3B
PD0325901	MAP2K1, MAP2K2	BRAF, MAP2K1	BRAF, MAP2K1, MAPK1	BRAF, MAP2K1, MAP2K2, MAPK1	BRAF, MAP2K1, MAP2K2, MAPK1
PI103	PIK3CA, PIK3CB, PIK3CD	PIK3CA, MTOR	PIK3CA, MTOR, MLST8	PIK3CA, MTOR, MLST8	BRAF, CHUK, FLT3, JAK1 LIMK2, MLST8, MTOR, PIK3CA, RAF1
PKF118-310	CTNNB1	-	-	DUSP1	CTNNB1, DUSP1, TCF7L2
JNK-IN-8	MAPK8, MAPK9, MAPK10	MAPK8, MAPK9	MAPK8, MAPK9	MAPK8, MAPK9	MAPK8, MAPK9
BI-D1870	RPS6KA1, RPS6KA3, RPS6KA2, RPS6KA6	RPS6KA1, RPS6KA3	PLK1, RPS6KA1, RPS6KA2, RPS6KA3	CSNK1A1, JAK2, PLK1, RPS6KA1, RPS6KA2, RPS6KA3	CSNK1A1, GSK3B, JAK2, PLK1, RPS6KA1, RPS6KA2, RPS6KA3
BI605906	IKBKB	-	-	IKBKB	IKBKB
Ruxolitinib	JAK1, JAK2	JAK1, JAK2	JAK1, JAK2, ROCK1	BMPR2, IRAK1, JAK1, JAK2, MAP2K3, MAP3K7, PLK1, ROCK1, RPS6KA1, RPS6KA2, RPS6KA4, RPS6KA5	BMPR2, BRAF, CSNK1A1, IKBKB, IRAK1, JAK1, JAK2, KIT, MAP2K1, MAP2K2, MAP2K3, MAP2K4, MAP3K7, MAPK8, PLK1, ROCK1, RPS6KA1, RPS6KA2, RPS6KA4, RPS6KA5
SB-505124	TGFBR1, ACVR1B, ACVR1C	-	ACVR1B, MAPK14, TGFBR1	ACVR1B, MAPK14, TGFBR1	ACVR1B, MAPK14, TGFBR1
D4476	CSNK1D, TGFBR1	CSNK1D, CSNK1E	CSNK1D, CSNK1E	CSNK1A1, CSNK1D, CSNK1E, MAPK14	CSNK1A1, CSNK1D, CSNK1E, MAPK14
10058-F4	MYC	-	-	-	MYC
Stattic	STAT3	-	-	-	STAT3
GSK2334470	PDPK1	PDPK1	PDPK1	PDPK1	PDPK1, ROCK1
PRT062607	SYK	SYK	PRKCA, SYK	FLT3, PRKCA, SRC, SYK	FLT3, JAK2, PRKCA, SRC, SYK

Appendix A.5 - Prediction Performance

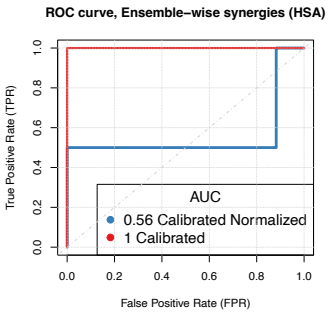
The quality of the synergy predictions obtained through simulations with the DrugLogics pipeline was assessed by plotting receiver operating characteristic (ROC) and precision-recall (PR) curves. Simulations were performed using CASCADE 1.0 and CASCADE 2.0, and the models were perturbed according to the target profiles shown in Tabel A.6 and Table A.7, respectively. The ROC and PR curves plotted for simulations with CASCADE 1.0 are shown in Figure A.2 and Figure A.3, respectively. Similarly, the curves plotted for simulations with CASCADE 1.0 are shown in Figure A.4 and Figure A.5, respectively.



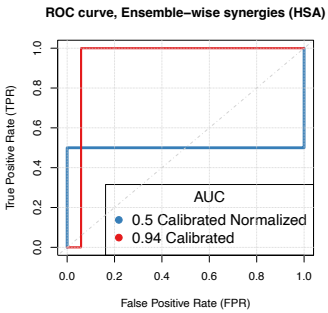
(a) Primary target



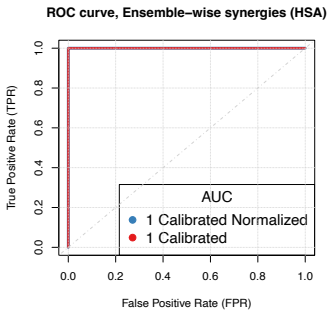
(b) 10 nM



(c) 100 nM

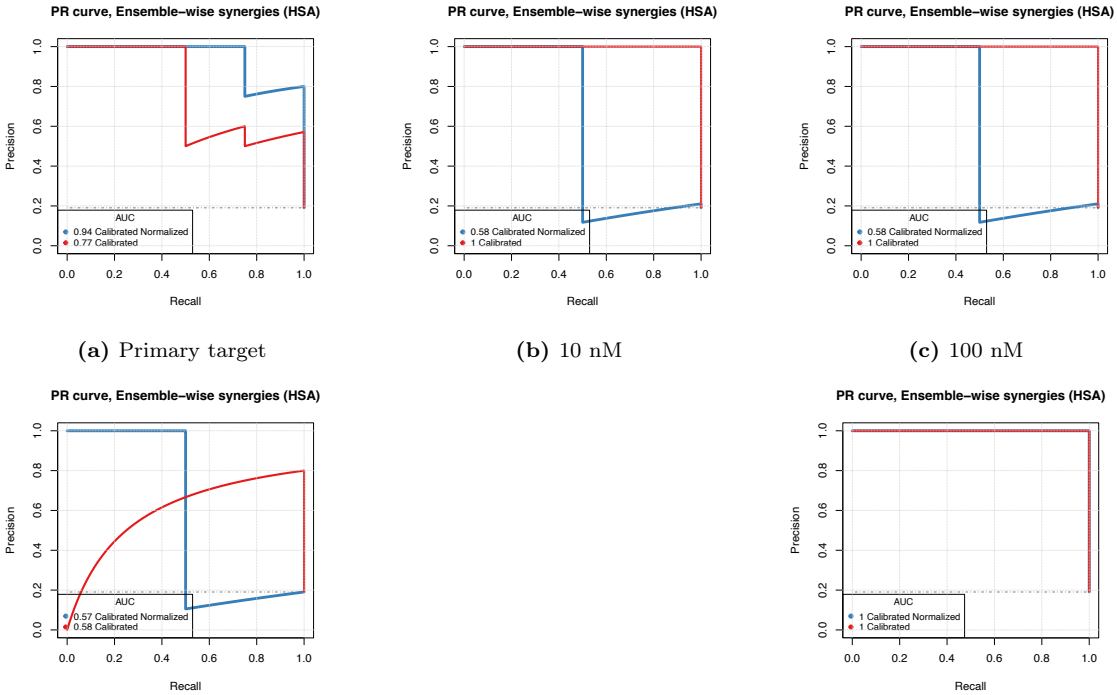


(d) 1 µM



(e) 10 µM

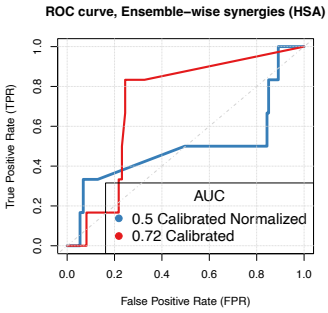
Figure A.2: Receiver operating characteristic (ROC) curves illustrating prediction performance achieved through modelling with CASCADE 1.0 model (HSA synergy method, 150 models calibrated to the adenocarcinoma cancer cell line (AGS) steady state represented by a red line, and normalised to a random proliferative phenotype, represented by a blue line). The models were perturbed according to a) primary target, and target profiles generated with DrugProfiler tool using binding affinity cut-offs equal to b) 10 nM, c) 100 nM, d) 1 µM, and e) 10 µM.



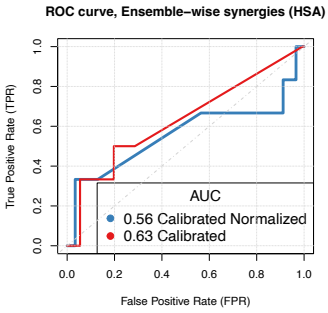
(a) Primary target **(b)** 10 nM **(c)** 100 nM

(d) 1 µM **(e)** 10 µM

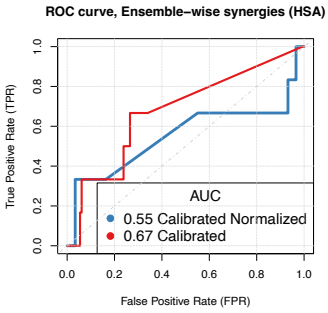
Figure A.3: Precision-recall (PR) curves illustrating prediction performance achieved through modelling with CASCADE 1.0 model (HSA synergy method, 150 models calibrated to the adenocarcinoma cancer cell line (AGS) steady state represented by a red line, and normalised to a random proliferative phenotype, represented by a blue line). The models were perturbed according to a) primary target, and target profiles generated with DrugProfiler tool using binding affinity cut-offs equal to b) 10 nM, c) 100 nM, d) 1 µM, and e) 10 µM.



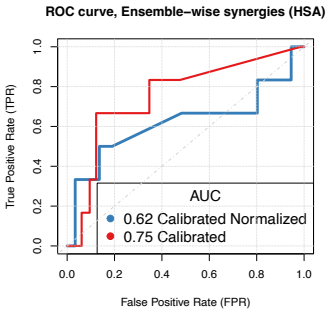
(a) Primary target



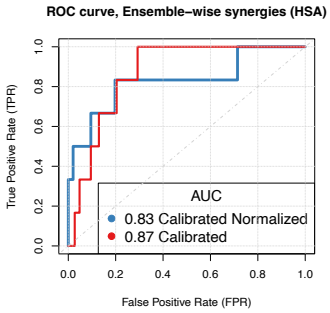
(b) 10 nM



(c) 100 nM

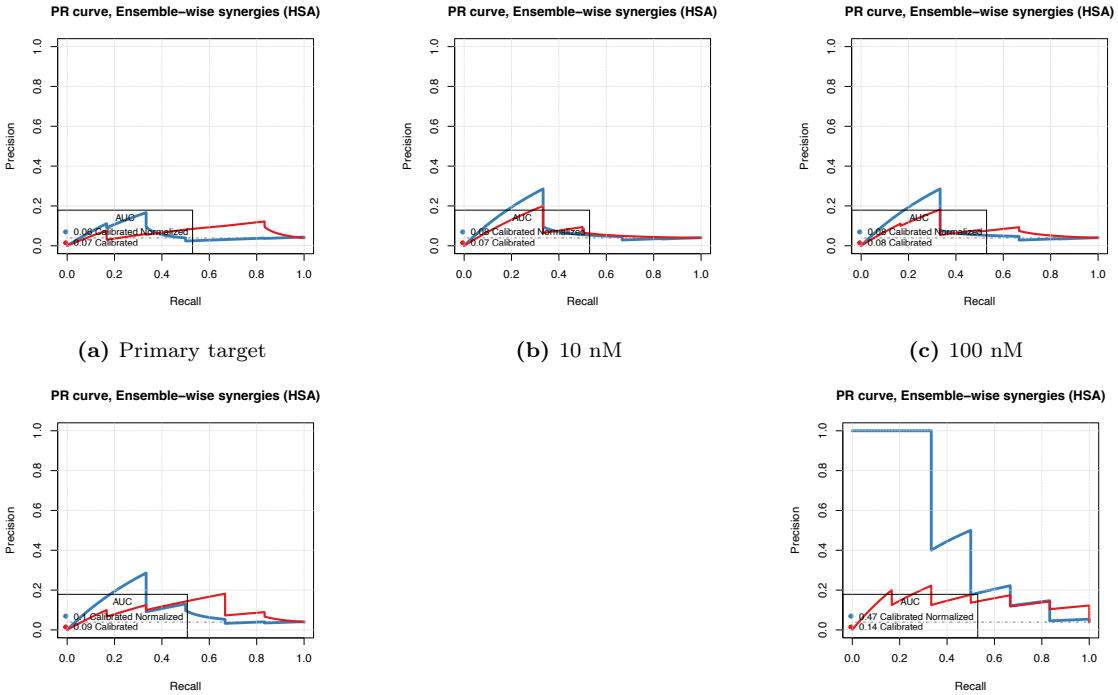


(d) 1 µM



(e) 10 µM

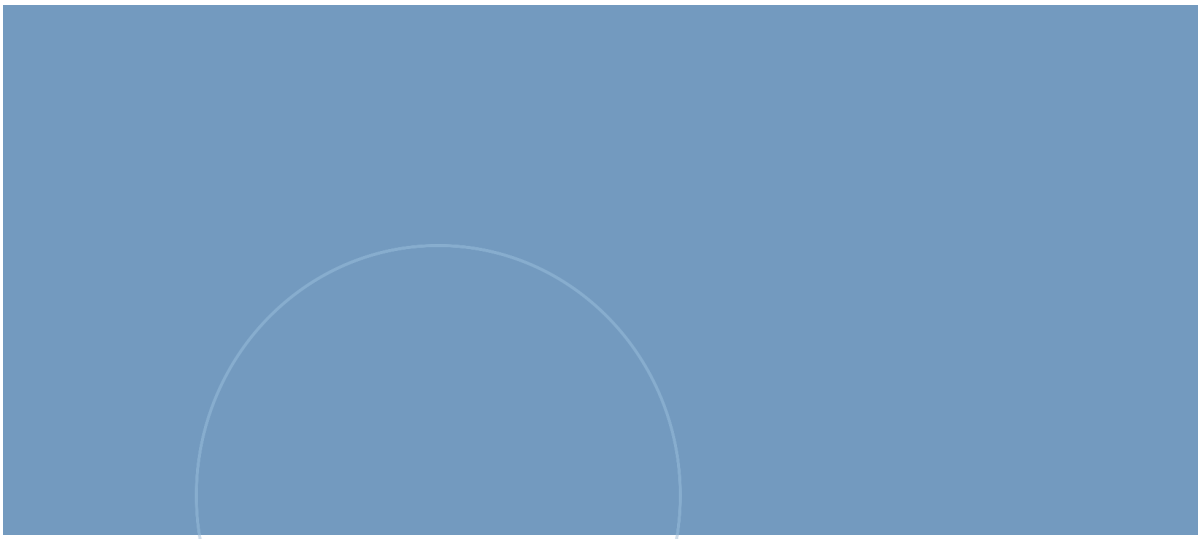
Figure A.4: Receiver operating characteristic (ROC) curves illustrating prediction performance achieved through modelling with CASCADE 2.0 model (HSA synergy method, 150 models calibrated to the adenocarcinoma cancer cell line (AGS) steady state represented by a red line, and normalised to a random proliferative phenotype, represented by a blue line). The models were perturbed according to a) primary target, and target profiles generated with DrugProfiler tool using binding affinity cut-offs equal to b) 10 nM, c) 100 nM, d) 1 µM, and e) 10 µM.



(a) Primary target **(b)** 10 nM **(c)** 100 nM

(d) 1 µM **(e)** 10 µM

Figure A.5: Precision-recall (PR) curves illustrating prediction performance achieved through modelling with CASCADE 2.0 (HSA synergy method, 150 models calibrated to the adenocarcinoma cancer cell line (AGS) steady state represented by a red line, and normalised to a random proliferative phenotype, represented by a blue line). The models were perturbed according to a) primary target, and target profiles generated with DrugProfiler tool using binding affinity cut-offs equal to b) 10 nM, c) 100 nM, d) 1 µM, and e) 10 µM.



NTNU
Norwegian University of
Science and Technology