



universität
wien

DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

„Structure-based analysis of targets involved in multiple myeloma“

verfasst von / submitted by

Melissa Leopold

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Magistra der Pharmazie (Mag.pharm.)

Wien, 2019 / Vienna, 2019

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

A 449

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Diplomstudium Pharmazie

Betreut von / Supervisor:

Univ.-Prof. Mag. Dr. Gerhard Ecker

Mitbetreut von / Co-Supervisor:

Dr. Claire Colas

Dedicated to my sister Martina

Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisor Univ.-Prof. Mag. Dr. Gerhard Ecker for giving me the opportunity to work on my thesis in his extraordinary group. I could not imagine a better supervisor – despite his permanently full calendar, he always found time for meetings and contributed with good advice how to drive the project forward. What I personally liked the most were his encouraging words and the fact that he believes in the potential of everyone. Thank you for the insight in scientific working!

My grateful thanks are also extended to my fantastic co-supervisor Dr. Claire Colas. From the very beginning, she guided me patiently through the whole computational work of my thesis and answered successively my never-ending list of questions. I appreciate her continuous commitment related to the research work, but also that she took her time and worked with me on my presentation skills. Thank you for your strong support and for being an inspirational role model.

I offer my sincere thanks also to Jakob Hager, who shared his data of the OpenPHACTS analysis and Giulia Banci for providing me her KNIME-workflow – it was really a big help for me and especially this work. Many thanks also to Dr. Riccardo Martini, who introduced me to LigandScout and helped me with the generation of the pharmacophore model.

Generally, I would like to thank the whole research group for a cheerful time in the lab, it was a pleasure to be part of the team and to learn from every one of you.

Last but not least, I would like to thank my family and friends – you were always there for me, in all ups and downs throughout my years of study. Thank you all so much!

TABLE OF CONTENT

I. ABSTRACT	- 8 -
II. ZUSAMMENFASSUNG	- 9 -
1 INTRODUCTION	- 10 -
1.1 BIOLOGICAL BACKGROUND OF MULTIPLE MYELOMA	- 11 -
1.1.1 DEFINITION	- 11 -
1.1.2 THE PHYSIOLOGICAL ROLE OF PLASMA CELLS IN THE IMMUNE SYSTEM	- 12 -
1.1.3 EPIDEMIOLOGY	- 13 -
1.1.4 PATHOGENESIS	- 14 -
1.1.5 DIAGNOSIS AND SYMPTOMS	- 16 -
1.1.6 TREATMENT	- 17 -
1.2 HIGH-RISK DRIVERS IN MULTIPLE MYELOMA	- 21 -
2 COMPUTATIONAL BACKGROUND	- 22 -
2.1 INTRODUCTION	- 22 -
2.2 COMPUTER-AIDED DRUG DESIGN	- 23 -
2.3 WORKFLOW FOR DISCOVERING NEW INHIBITORS	- 24 -
2.4 PHARMACOPHORE MODELING	- 25 -
2.4.1 OVERVIEW	- 25 -
2.4.2 STRUCTURE-BASED APPROACH	- 26 -
2.4.3 LIGAND-BASED APPROACH	- 27 -
2.5 MOLECULAR DOCKING	- 27 -
2.5.1 GENERAL INFORMATION	- 27 -
2.5.2 SEARCH STRATEGY	- 27 -
2.5.3 SCORING FUNCTIONS	- 28 -
2.6 BINDING AFFINITY	- 30 -
2.6.1 THERMODYNAMICAL BACKGROUND	- 30 -
3 AIM OF THE THESIS	- 32 -
4 MATERIAL AND METHODS	- 34 -
4.1 DATA COLLECTION	- 34 -
4.1.1 OPENPHACTS	- 34 -
4.1.2 PROTEIN DATA BANK	- 34 -
4.1.3 ChEMBL	- 34 -
4.1.4 PUBMED	- 34 -
4.2 PHARMACOPHORE GENERATION	- 35 -
4.3 LIGAND DOCKING	- 36 -
4.3.1 PROTEIN PREPARATION AND GRID GENERATION	- 37 -
4.3.2 LIGAND PREPARATION	- 37 -
4.3.3 DOCKING	- 37 -
4.3.4 REDOCKING PKMYT1	- 38 -

4.4	BINDING AFFINITY CALCULATIONS	- 38 -
5	RESULTS AND DISCUSSION	- 40 -
5.1	PKMYT1	- 40 -
5.2	PHARMACOPHORE VALIDATION, REDOCKING AND VIRTUAL SCREENING	- 43 -
5.3	COMPOUND SELECTION	- 46 -
5.3.1	SELECTED COMPOUNDS FOR IN VITRO TESTING	- 49 -
6	CONCLUSION	- 52 -
7	BIBLIOGRAPHY	- 53 -
8	APPENDIX	- 56 -
8.1	LIST OF ABBREVIATIONS	- 56 -
8.2	LIST OF FIGURES	- 57 -
8.3	LIST OF TABLES	- 58 -

I. Abstract

Multiple myeloma is a malignant disease of the hematopoietic system. The cancerous disease is characterized by an uncontrolled proliferation of plasma cells, an important antibody-producing cell type of the immune system. Until this day, multiple myeloma is treatable, but still incurable. Surviving rates range from a few months to more than ten years, depending on myeloma cell biology and individual risk factors of each patient.¹ Previous work on this topic revealed a pathway of seventeen target genes involved in a high-risk multiple myeloma, which is defined by fast progression or death within 1,5 years.²

The aim of this thesis was a structure-based analysis of formerly mentioned targets by means of computational methods. Our main focus was on PKMYT1, a kinase involved in cell cycle regulation. With the help of an X-ray structure of PKMYT1 complexed with a known inhibitor from the Protein Data Bank, we generated a pharmacophore model and used it to screen a manually curated database of approved, withdrawn, and experimental drugs of the platform Drugbank, in order to find new inhibitors.

The output of this screening was docked into the PKMYT1 protein structure and we were able to predict eight compounds as potential inhibitors for our target. All of the final compounds are experimental drugs, which means they are preclinically tested. A follow-up in-vitro test of the compounds is still pending.

II. Zusammenfassung

Das multiple Myelom ist eine bösartige Erkrankung des blutbildenden Systems. Diese Krebserkrankung ist gekennzeichnet durch eine unkontrollierte Vermehrung von Plasmazellen, die wichtige Antikörper-produzierende Zellen des Immunsystems darstellen. Bis zum heutigen Tag ist das multiple Myelom zwar behandelbar, aber immer noch unheilbar. Die Überlebensraten reichen von ein paar Monaten bis zu mehr als 10 Jahren, abhängig von der Zellbiologie des Myeloms und individuellen Risikofaktoren jedes Patienten.

Eine vorhergehende Arbeit zu diesem Thema offenbarte ein neues Netzwerk-Modell von 17 Proteinen, die in ein Hochrisiko-Myelom involviert sind, das sich durch Fortschreiten der Erkrankung oder Tod des Patienten innerhalb von 1,5 Jahren auszeichnet.

Das Ziel dieser Arbeit war eine struktur-basierte Analyse von den erwähnten 17 Zielstrukturen mit computergestützten Methoden. Unser Hauptaugenmerk lag auf PKMYT1, einer Kinase, die an der Zellzyklus-Regulation teilnimmt. Mithilfe einer Röntgenstruktur von PKMYT1 in Komplex mit einem bereits bekannten Inhibitor aus der Proteindatenbank (PDB) generierten wir ein Pharmakophormodell und benutzten es, um eine manuell erstellte Datenbank aus bereits zugelassenen, zurückgezogenen und experimentellen Wirkstoffen zu durchsuchen, um dadurch neue Inhibitoren zu finden.

Die resultierenden Verbindungen der Datenbanksuche wurden zurück in die Proteinstruktur von PKMYT1 gedockt und wir waren in der Lage, acht der Strukturen als potentielle Inhibitoren für unser Protein vorherzusagen. Alle finalen Verbindungen sind experimentelle Wirkstoffe, das heißt, sie wurden nur präklinisch getestet. Ein anschließender in-vitro Test zur Überprüfung unserer Vorhersage steht noch aus.

1 Introduction

Nowadays cancer is one of the leading diseases causing death worldwide. It is a term for a large group of related malignant diseases. All of them have in common the development of abnormal cells which grow out of control, divide rapidly and are able to invade tissues or organs throughout the whole body.³ Our work focuses on one specific type of cancer, the multiple myeloma. It is a representative of hematological cancers and primarily affects plasma cells.

Although many new therapeutic strategies evolved in the past years, multiple myeloma still remains an incurable disease. The patients survival strongly depends on the course of the disease, including also the host factors, tumor burden, cytogenetic abnormalities and response to therapy.⁴ Generally, survival ranges from a few months to more than 10 years⁵, with the median survival of 5-7 years.⁴ Patients with high risk factors, such as certain cytogenetic abnormalities are identified with poor prognosis.⁶ Although multiple myeloma is a rare type of cancer, prevalence rates are rising. For example, the increasing age of population, new diagnostic techniques and improved patient survival are contributing factors.⁷

As already mentioned, in the past decades therapy improved due to the establishment of new drugs on the market. Immune modulatory drugs (lenalidomide, thalidomide) and proteasome inhibitors (bortezomib) were identified as highly effective in the treatment of multiple myeloma and replaced older drugs like alkylators, anthracyclines or interferons.⁸ Actually, also several promising drug candidates are in phase II and III studies and will presumably be approved in the near future.

The purpose of our work was to find compounds, which could be used in the therapy of a high-risk multiple myeloma. But before switching to the aim of this thesis, I want to explain the biological background of the disease and actual therapy approaches more in detail.

1.1 Biological Background of Multiple Myeloma

1.1.1 Definition

The multiple myeloma is defined by the National Cancer Institute of America (NCI) as a plasma cell neoplasm and belongs to cancer diseases of the hematopoietic system. Such neoplasms, which are characterized by an excessively growth of tissue, emerge when abnormal plasma cells (also known as multiple myeloma cells) form malignant tumors in bones or soft tissue.⁹

For a better understanding, the following image explains the development of blood stem cells into plasma cells.

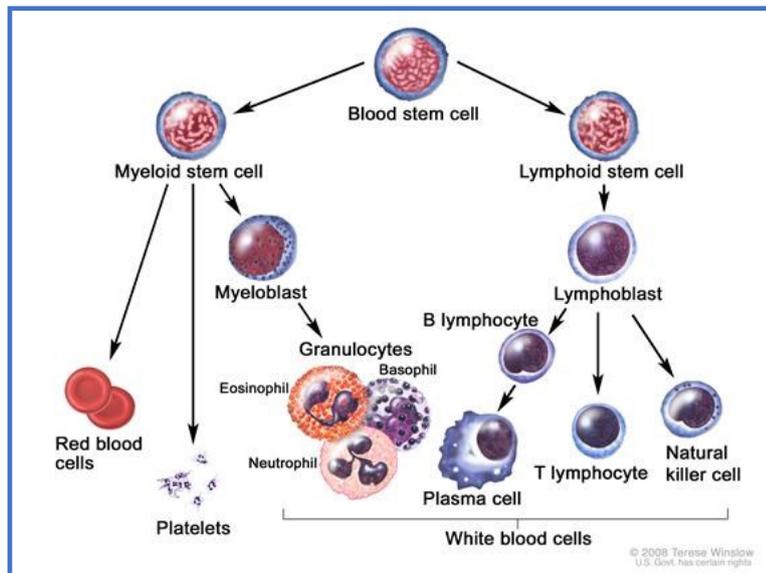


Figure 1: Differentiation of blood stem cells. Source: <https://www.cancer.gov/images/cdr/live/CDR596562-750.jpg>

In Figure 1, the pathway of a physiological differentiation of blood stem cells is depicted. Stem cells, which are produced in the bone marrow, can either turn into myeloid or lymphoid stem cells. From myeloid stem cells can result red blood cells, platelets or granulocytes. Lymphoid stem cells develop into lymphoblasts and finally result in natural killer cells, T-lymphocytes or B-lymphocytes. The latter B-lymphocytes (or B-cells) undergo several processes and mature into plasma cells.

White blood cells, consisting of granulocytes, plasma cells, T-lymphocytes and natural killer cells, are important members of the immune system. The function of these cells is to get rid of invading pathogens and protect the body against diseases.

1.1.2 The physiological role of plasma cells in the immune system

To gain a better insight on the impact of plasma cells, it is relevant to have a short overview of the functions and cells of the immune system.

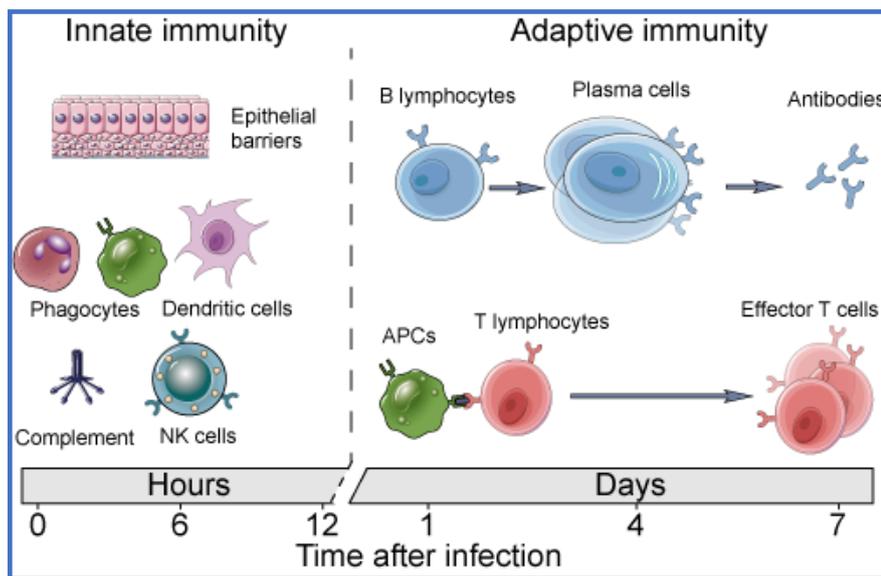


Figure 2: The innate and acquired immune system. Source: <https://www.creative-diagnostics.com/innate-and-adaptive-immunity.htm>

The complex immune system is generally composed of the innate and the acquired part and its common function is to protect organisms against diseases. The innate immune system consists of physical epithelial barriers, phagocytic leukocytes, dendritic cells, and natural killer (NK) cells.¹⁰ However, it is not specific and responds to diverse pathogens in the same manner, usually at first through macrophages and neutrophils.¹¹

The acquired immune system works in a different way. On the one hand, killer T-cells and helper T-cells recognize antigenic peptides only when they are bound to HLA (= human leukocyte-associated antigens) molecules.¹² In detail, killer T-cells recognize antigens on class I HLA molecules and helper T-cells recognize antigens represented by class II HLA proteins. Killer T-cells are able to destroy virus- or parasite-infected

cells immediately, whereas helper T-cells activate B-cells and “help” thereby to generate antibodies.¹¹

On the other hand, B-lymphocytes possess B-cell receptors (which are actually antibodies) on their cell membrane, allowing them to bind specific antigens. As soon as an antigen is bound, the complex gets internalized and processed antigens are exposed on HLA-molecules on the cell’s surface to attract helper T-cells. Helper T-cells activate B-cells, by permitting them to mature into either memory B-cells or plasma cells. While plasma cells are able to actively secrete millions of antibodies (= immunoglobulins) against their antigen, memory B-cells persist for years in the human body, only producing antibodies if they are reactivated by a confrontation with their matching antigen.¹¹ Antibodies are efficient in an early defense regarding reinfections, are able to bind pathogens and neutralize toxins.¹¹

1.1.3 Epidemiology

In total, multiple myeloma is accounting for 10% of all hematological malignancies and on third place ranked behind lymphoma and leukemia.¹³ The disease mainly concerns elderly people with a median age of diagnosis at 66 years.¹⁴ Furthermore, the incidence rate varies worldwide and is highest in developed countries like the United States, Europe and Australia.⁷ Although it is a relative rare type of cancer, there is an expectation of about 32,110 new cases diagnosed and about 12,960 deaths to occur in 2019 in the United States.¹⁵

As a result of better diagnostic techniques and an advanced patient survival, caused by improved therapies such as the autologous hematopoietic stem cell transplantation (ASCT) or novel therapeutic agents, prevalence rates are successively increasing.¹⁶ However, despite the advances in therapy, it is still incurable.

1.1.4 Pathogenesis

The multiple myeloma is characterized by an uncontrolled proliferation of abnormal plasma cells, which produce so called monoclonal protein (M-protein), a monoclonal immunoglobulin or immunoglobulin light chain. As the cells derive from bone marrow, they can form tumors in several bones of the body.

Physiologically, B-cells, which develop from hematopoietic stem cells in the bone marrow, migrate to secondary lymphoid organs like the lymph node or spleen and are continuously processed until they become a mature plasma cell.⁷ Such plasma cells produce in healthy human immunoglobulins (or antibodies) to fight infections. (see Figure 3)

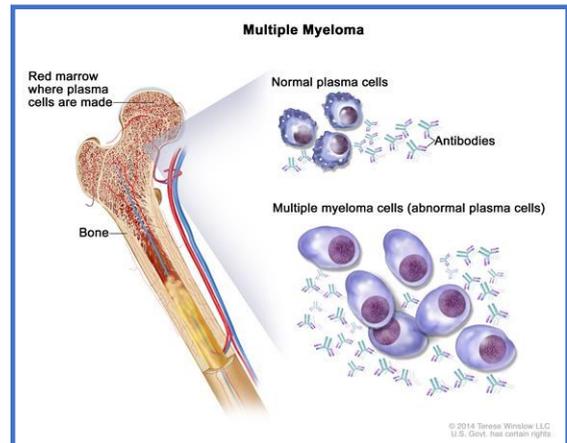


Figure 3: Difference between normal plasma cells and malignant plasma cells (or myeloma cells). Normal plasma cells produce immunoglobulins (=antibodies). Contrarily, the multiple myeloma cells that are bigger in size and unable to produce effective antibodies. In fact, they produce M-protein, a sort of abnormal immunoglobulin fragment which leads to impaired function of the immune system, blood thickness and kidney damage. (Picture reproduced from: <https://www.cancer.gov/images/cdr/live/CDR763079-750.jpg>)

In the multiple myeloma, chromosomes and genes are damaged, which leads to a production of abnormal plasma cells and a following uncontrolled secretion of M-protein – mostly functionless monoclonal antibodies or parts of antibodies. Moreover, high levels of M-protein can lead to an impaired function of the immune system, to blood thickness and kidney damage.

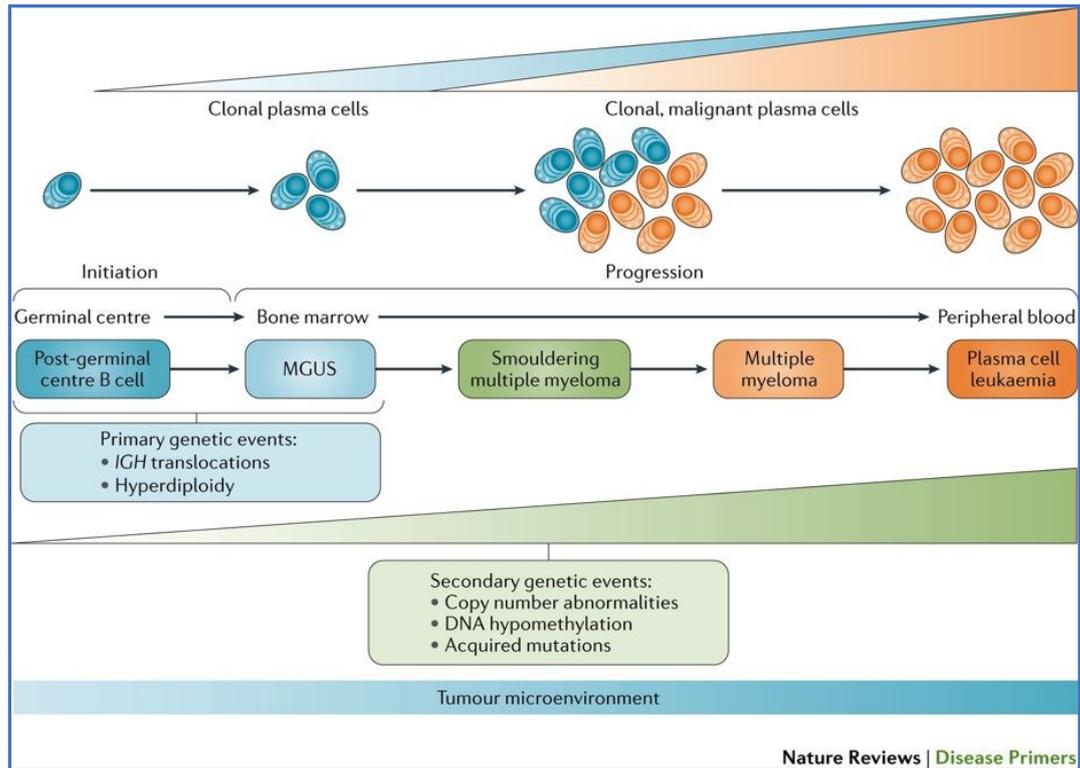


Figure 4: Overview of the development and previous stages of the multiple myeloma. (Picture reproduced from Kumar, S. K. et al. Multiple myeloma, Nature Reviews Disease Primers 3, 2017)

The development of the disease comprises a multi-step process. In particular, initiating events like chromosomal translocations, e.g. involving IGH (genes that encode immunoglobulin heavy chains) or aneuploidy promote an increasing proliferation of plasma cells.⁷ Additionally, secondary genetic events like mutations and epigenetic alterations, e.g. DNA hypomethylation might also contribute to a development of multiple myeloma.⁷

Primary genetic events can lead to a clinical manifestation of a monoclonal gammopathy of undetermined significance (MGUS). MGUS is a pre-malignant state, where patients do not suffer from myeloma-typical symptoms (e.g. like bone damage, excessive plasma cells in the marrow or anemia), but a low level of M-protein is already determined in the blood. If the M-protein level raises and plasma cells in the bone marrow increase, the disease develops into a smouldering multiple myeloma (SMM), which represents a transitional state between MGUS and multiple myeloma (MM). Patients with those particular preceding forms can develop a multiple myeloma eventually. Another very aggressive form of the multiple myeloma is the plasma cell leukemia, where a high level of malignant plasma cells can be determined in the peripheral blood.

1.1.5 Diagnosis and Symptoms

The International Myeloma Working Group (IMWG) published in 2014 criteria that must be met for a differential diagnosis of multiple myeloma and related disorders.¹⁷ The criteria of the three most important forms, monoclonal gammopathy of undetermined significance (MGUS), smouldering multiple myeloma (SMM), and multiple myeloma (MM), are displayed in the following table:

Differential diagnosis of multiple myeloma and related diseases

Disorder:	MGUS (monoclonal gammopathy of undetermined significance)	SMM (smouldering multiple myeloma)	MM (multiple myeloma)
	All 3 criteria must be met:	Both criteria must be met:	Both criteria must be met:
	<ul style="list-style-type: none"> Serum monoclonal protein (non-IgM type) <3gm/dL 	<ul style="list-style-type: none"> Serum monoclonal protein (IgG or IgA) ≥3gm/dL, or urinary monoclonal protein ≥500 mg per 24h and/or clonal bone marrow plasma cells 10–60% 	<ul style="list-style-type: none"> Clonal bone marrow plasma cells ≥10% or biopsy-proven bony or extramedullary plasmacytoma
	<ul style="list-style-type: none"> Clonal bone marrow plasma cells <10%* 	<ul style="list-style-type: none"> Absence of myeloma defining events or amyloidosis 	<ul style="list-style-type: none"> Any one or more of the following myeloma defining events: <ul style="list-style-type: none"> Evidence of end organ damage that can be attributed to the underlying plasma cell proliferative disorder, specifically: <ul style="list-style-type: none"> Hypercalcemia: serum calcium >0.25 mmol/L (>1 mg/dL) higher than the upper limit of normal or >2.75 mmol/L (>11 mg/dL) Renal insufficiency: creatinine clearance <40 mL per minute or serum creatinine >177 μmol/L (>2 mg/dL) Anemia: hemoglobin value of >2 g/dL below the lower limit of normal, or a hemoglobin value <10 g/dL Bone lesions: one or more osteolytic lesions on skeletal radiography, computed tomography (CT), or positron emission tomography-CT (PET-CT) Clonal bone marrow plasma cell percentage ≥60% Involved: uninvolved serum free light chain (FLC) ratio ≥100 (involved free light chain level must be ≥100 mg/L) >1 focal lesions on magnetic resonance imaging (MRI) studies (at least 5mm in size)
	<ul style="list-style-type: none"> Absence of end-organ damage such as hypercalcemia, renal insufficiency, anemia, and bone lesions (CRAB) that can be attributed to the plasma cell proliferative disorder 		

Figure 5: Differential diagnosis criteria for monoclonal gammopathy of undetermined significance, smouldering multiple myeloma and multiple myeloma. Criteria reproduced from Rajkumar SV, Kumar S. Multiple Myeloma: Diagnosis and Treatment. Mayo Clin Proc. 2016;91(1):101-19.

The guidelines are based on the measurement of serum monoclonal protein level, bone marrow infiltration of plasma cells and myeloma-defining events, the latter one being a summary of CRAB-criteria and new biomarkers.

CRAB is an acronym for **calcium** (hypercalcemia), **renal** insufficiency, **anemia** and **bone** lesions, which are all typical signs of multiple myeloma.

To the most frequent symptoms of the multiple myeloma count bone pain and fatigue. Anemia appears in 75% of the patients and is one of the contributing factors to fatigue.⁴ Bone pain is caused by myeloma cells, which produce receptor activator for nuclear factor κ B ligand (RANKL) and activate thereby osteoclasts, a bone-resorbing cell type.⁵ Increased osteoclast activity leads to lytic bone lesions, bone destruction and

hypercalcemia.⁵ Most commonly the pain occurs in the back, hips or skull of patients. Another accessory symptom is osteoporosis, where bones lose their stability and the risk for fractures increases.

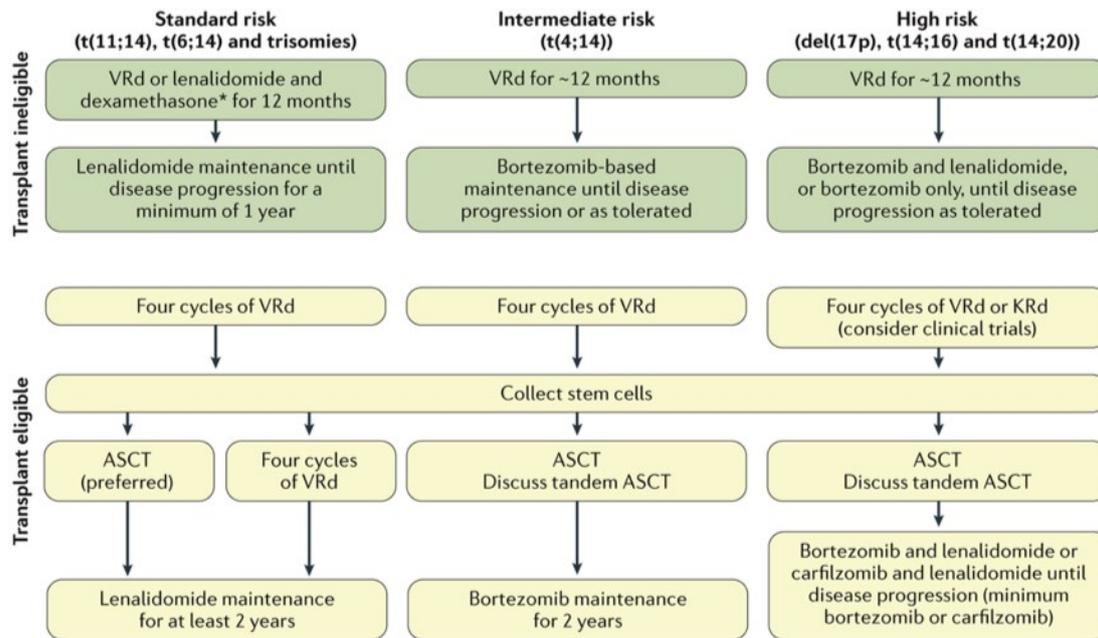
Furthermore, neurological symptoms can appear due to hypercalcemia or spinal cord depression.

1.1.6 Treatment

The treatment of multiple myeloma changed over the past decades due to a gain of knowledge in the disease pathology and improvements in therapeutic drug classes.

A therapy strongly depends on the course of the disease and involves pharmacological interventions and other clinical methods like an ASCT (= autologous stem cell transplantation) to control the progress.

The following figure gives an overview of the current suggested therapy options, subdivided in standard/intermediate/high-risk patients and transplant eligible/ineligible patients.⁷



Nature Reviews | Disease Primers

Figure 6: Therapeutic scheme for multiple myeloma, subdivided into standard risk, intermediate risk and high-risk patients and autologous hematopoietic stem cell transplantation (ASCT) eligibility or ineligibility. VRd (= bortezomib + lenalidomide + dexamethasone); KRd (= carfilzomib, lenalidomide, dexamethasone); ASCT (= autologous stem cell transplantation) (Image reproduced from Kumar, S. K. et al. Multiple myeloma, Nature Reviews Disease Primers 3 2017)

A therapy starts for every patient preferably with the combination of a proteasome inhibitor and an immunomodulatory drug. Afterwards, therapy options split for transplant eligible and ineligible patients, so it has to be considered if the patient can receive a stem cell transplantation or not. An ASCT comprises the removal of stem cells out of the peripheral blood of the patient, the storage of it, a process called myeloablative conditioning which consists of chemotherapy and/or radiation to destroy myeloid cells and a final reinfusion of the collected stem cells. Most studies limit an ASCT conduction to patients ≤ 65 years of age.¹⁸ However, sometimes the physiological age weighs more than the chronological age, therefore exceptions can be made if the patient feels fit enough.¹⁹ Another important determinant to decide for an ASCT are comorbidities. Patients, that suffer from severe cardiac or pulmonary diseases should not be offered a transplantation.⁷ Nevertheless, studies came to the conclusion that also in times of high efficacy drugs, ASCT still remains a solid method to enhance responses and to improve overall and progression-free survival rates in multiple myeloma patients.²⁰ Next, patients are subdivided into a standard-, intermediate- or high-risk category, depending mainly on their genetical abnormalities. However, as initial therapy for all pa-

tients able to tolerate a multidrug therapy the VRD regimen is applied. VRD is a combination of the proteasome inhibitor bortezomib, the immunomodulatory drug lenalidomide and the corticoid drug dexamethasone. Proteasome inhibitors in combination with immunomodulatory drugs are currently the best therapeutic option for patients with a newly diagnosed multiple myeloma.⁷ Specifically, the VRD regimen showed an improved progression-free survival and overall survival compared to patients, which were treated with lenalidomide and dexamethasone only.²¹

Patients ineligible for a stem cell transplantation, the VRD regimen or a regimen based on lenalidomide and dexamethasone only is the preferred choice. Also, treatment has always to be modified on the patient's characteristics, including age, performance and frailty.⁷

Since the multiple myeloma is stated as an incurable disease, therapy approaches aim to reduce the progress of the disease and prolong the surviving rate as well as to reduce the underlying symptoms like bone disease, pain or anemia to a minimum.

In case of bone lesions, bisphosphonates are therapy of first choice, as they can delay the development of bone lesions and prevent fractures.²² Pain treatment should start with non-opioid analgesics and avoid non-steroidal anti-inflammatory drugs (NSAIDs), because they can lead to renal failure. If therapy with non-opioid drugs is not sufficient, opioid analgesics can be supplemented to achieve an optimal pain relief.⁷ Furthermore, transfusions with erythropoiesis-stimulating agents, iron and vitamin support is recommended to relief anemia respectively fatigue.²²

Overview of important drug classes currently used in therapy of multiple myeloma

Proteasome Inhibitors	Immunomodulatory drugs	Monoclonal antibodies	Histone deacetylase inhibitor	Alkylating drugs	Corticoid drugs	Other drugs
Bortezomib	Lenalidomide	Daratumumab	Panbinostat	Melphalan	Dexamethasone	Cisplatin
Carfilzomib	Thalidomide	Elotuzumab		Cyclophosphamide	Prednisone	Etoposid
Ixazomib	Pomalidomide			Bendamustine		Doxorubicin

Figure 7: Drug-classes and drugs used in therapy of multiple myeloma.

Investigational drugs for multiple myeloma

Proteasome inhibitors	Monoclonal antibody	Other drugs
Marizomib	Isatuximab	Dinaciclib
Oprozomib		Filanesib
		LGH-447
		ABT-199 (venetoclax)

Figure 8: Investigational drugs for multiple myeloma.

1.2 High-risk drivers in multiple myeloma

We started our project based on a network model of our collaboration partners around Univ. Prof. Dr. Heinz Ludwig, head of Wilhelminen Cancer Research Institute in Vienna. The network model consists of target genes involved in a high-risk multiple myeloma. A high-risk status is characterized by progression of the disease or death within 18 months.²

Their statistical model (Figure 9) is based on a dataset of 645 patients from the IA9 MMRF CoMMpass trial, one of the largest long-term genomic research studies in multiple myeloma. The results of their work revealed 17 target genes identified as drivers of a high-risk disease. Subsequently, they tested their model on another dataset of the IFM/DFCI trial (a study about the relevance of ASCT in the era of new drugs) and performed experimental validation of the main molecular drivers using myeloma cell lines and small molecule inhibitors of MELK, CDK1 and PLK4.² This network model was published in order to help researchers to concentrate on the most promising targets responsible for high-risk multiple myeloma.

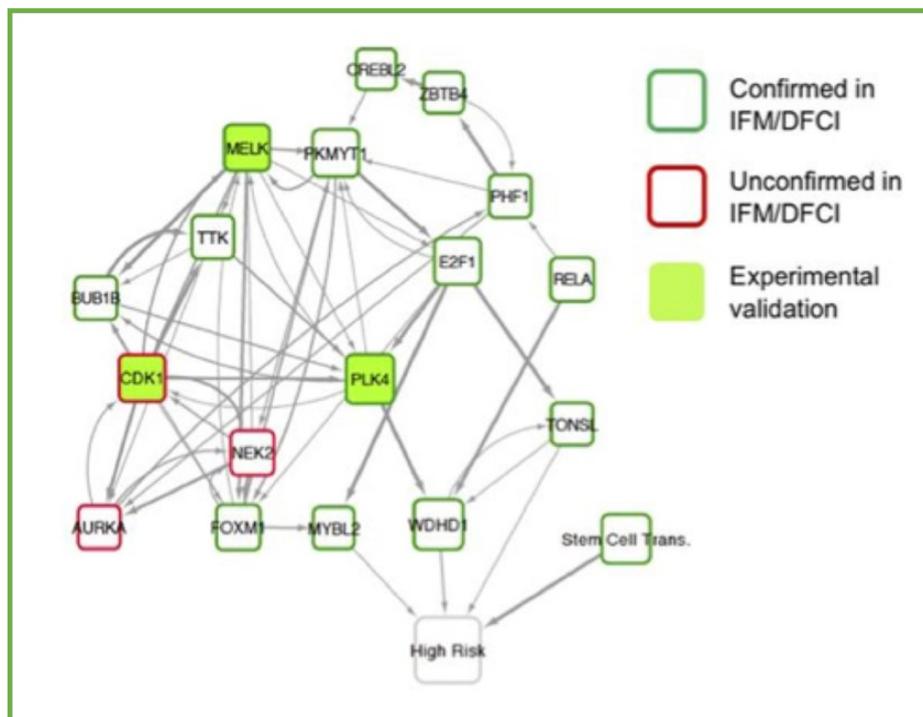


Figure 9: Network model displaying target genes involved in high-risk multiple myeloma. Picture reproduced from Furchtgott, L. et al. Multiple Myeloma Drivers of High Risk and Response to Stem Cell Transplantation Identified By Causal Machine Learning: Out-of-Cohort and Experimental Validation. *Blood* **130**, 3029–3029 (2017)

2 Computational Background

2.1 Introduction

Discovering a new drug and bringing it on the market is a very challenging process, since the development is very time-consuming and expensive. Throughout the process, it is never sure if the drug-candidate will be successfully approved by authorities. In history, people often discovered new active substances serendipitously. Nowadays more sophisticated options are available. Two important of them are High Throughput Screening (HTS) and Computer-Aided Drug Design (CADD). While HTS is an *in vitro* method enabling to assay thousands of substances on cells by robotic automations, Computer-Aided Drug Design works *in silico* (Figure 10). Figure 10 illustrates how CADD is implemented in the drug discovery process nowadays and also depicts different computational methods used in pharmaceutical research. The foundation of this process is the identification of the target. As soon as the molecular mechanisms that lead to a disease are identified and the target is validated, the drug discovery process can start. It is possible to use structure-based or ligand-based methods. Ideally, multiple lead compounds result from an effective CADD campaign. Then, the pharmacokinetic parameters (i.e. potency, efficacy, permeability and absorption) are optimized. Finally, successfully improved lead structures are tested *in vivo* to find drug candidates. After deciding for the final drug candidate, clinical trials can start.

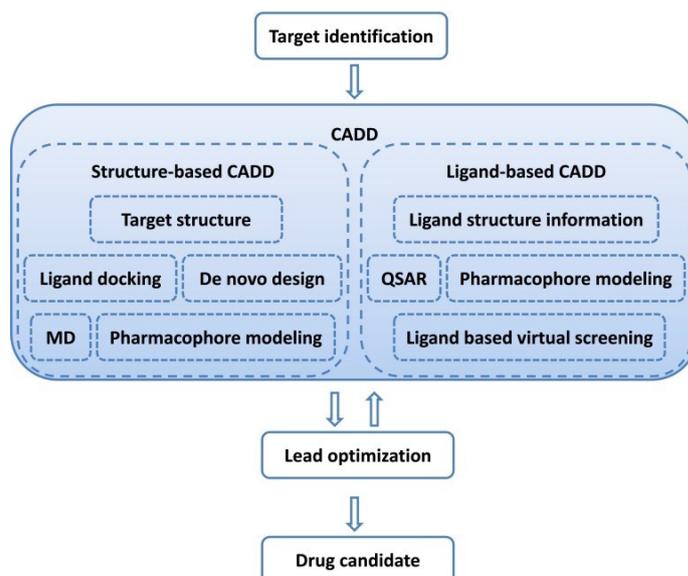


Figure 10: Drug discovery process involving computational techniques. Picture reproduced from Gregory Sliwoski et al. (Computational methods in Drug Discovery, 2013)

2.2 Computer-Aided Drug Design

In the recent years, computational methods gained attention and improved both drug discovery and research. The reasons why they are so popular today is that they save time and costs compared to high throughput screens, while yielding good results. Additionally, the concept of discovering a new drug is different: in HTS substances are tested on proteins or cells and the assay states if they are active at a certain concentration or not. In computational approaches, a drug is developed with focus on the molecular mechanism that leads to a certain disease.²³

The Computer-aided drug design can be generally separated into two main categories: the structure-based drug design and the ligand-based drug design. On the one hand, the structure-based design requires a three-dimensional structure of the protein, preferably in a high resolution. On the other hand, ligand-based design can be used when the structure of the target is unknown, but already several inhibitors are established, which can provide information about the relationship of structure and bioactivity.

In our work, only structure-based methods were implemented. They are as mentioned before based on the knowledge of the three-dimensional structure of a target. Via methods like X-ray diffraction, nuclear magnetic resonance spectroscopy or also elec-

tron microscopy, the three-dimensional positions of the atoms in a structure are detected and are further visualized on the computer. If a target structure is not available, there is the possibility to generate a homology model by using related proteins of a similar sequence, of which the structure was already identified.

Then, in the structure-based drug design workflows used can be distinguished between virtual screening and de novo design. In a virtual screening, a whole library of compounds, fragments or also lead-like structures are screened against a target in order to find the most valuable candidates for a following experimental testing. This screening should help to reduce the great number of structures to only a few candidates, which show the highest potential of having drug-like properties. Virtual screening is either used to identify new molecules for a specific target, or it can also be helpful in a repurposing attempt when it comes to patent existing molecules for another indication.²³

In the de novo design however, the three-dimensional structure is used to design a new molecule step by step into the binding pocket. This has the advantage to be innovative, since the new structure has never been synthesized before.²³

2.3 Workflow for discovering new inhibitors

In our case, an X-ray structure of the protein was available, and we used the following workflow presented in Figure 11 to find new inhibitors for our target.

At the stage of choosing a PDB-structure of the target, our criteria were a high resolution, and a structure in complex with an already known inhibitor. Based on the selected protein-inhibitor complex we generated a pharmacophore model, which is presented in detail in chapter 4 'Material and Methods'. Furthermore, we validated our model by setting up two databases with the biological activity data of ChEMBL, containing known actives and inactives. Our model was run on both databases. The results of this validation are discussed in the chapter 5 'Results and Discussion'. With our final model, we screened a database (Drugbank). Finally, we conducted a docking and selected the best compounds for a possible experimental assay.

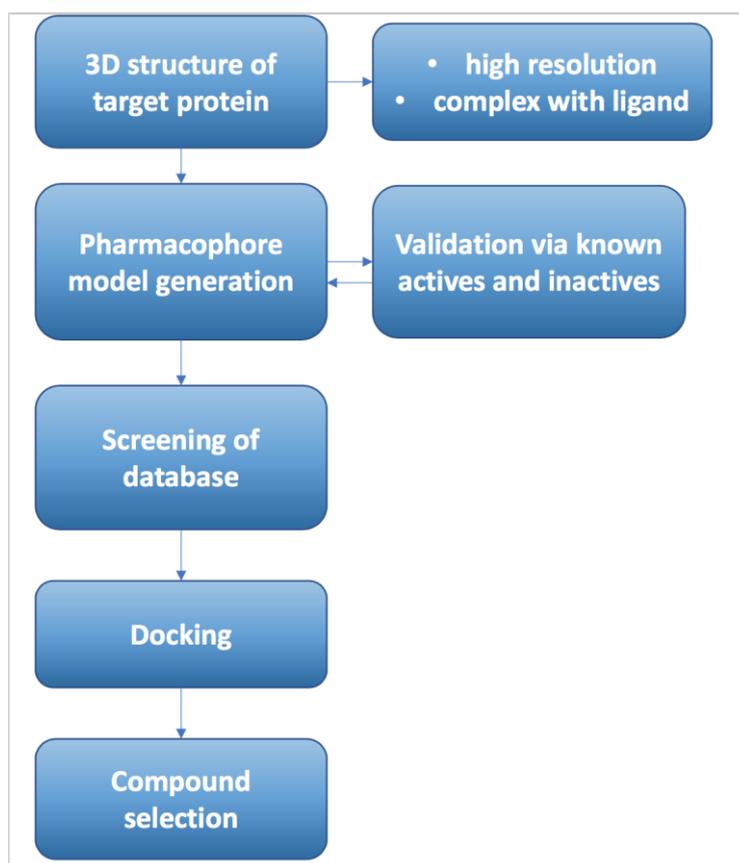


Figure 11: Our workflow to find new inhibitors.

2.4 Pharmacophore Modeling

2.4.1 Overview

IUPAC, the International Union of Pure and Applied Chemistry, defines pharmacophores as “an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response”.²⁴ In computational chemistry, a pharmacophore model describes features of molecules in a special geometric distance to each other. Features of active molecules can include hydrophobic parts, hydrogen bond acceptors and donors, positive and negative ionizable areas and aromatic systems. An example of a pharmacophore model and all possible molecule-describing features of LigandScout²⁵ are illustrated in Figure 12A and 12B.

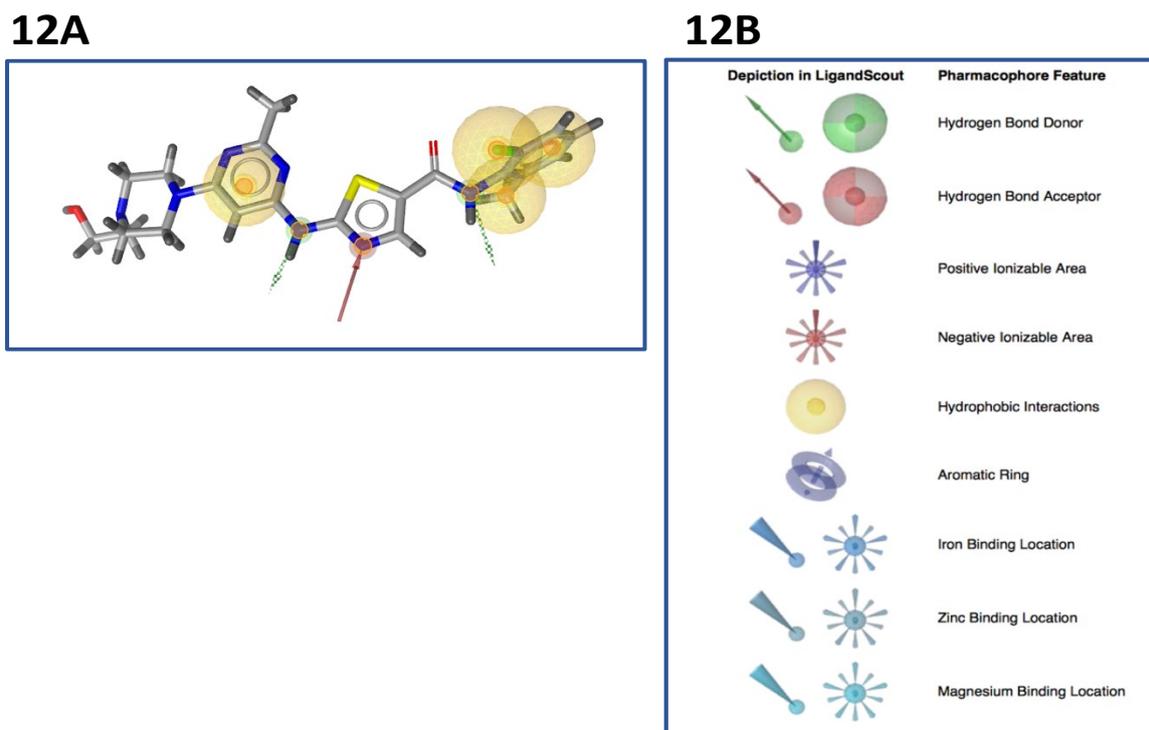


Figure 12: (A) Pharmacophore model based on kinase inhibitor dasatinib. Generated with LigandScout. (B) Feature explanations for LigandScout.

A pharmacophore model describes the interactions that can be formed between ligand and receptor. The functional groups and scaffolds behind can differ for every ligand matching the same model, while interactions with the protein stay the same. Consequently, it is assumed that also the biological activity will not change for the molecules that match the same pharmacophore.

The concept of pharmacophores is used to screen large databases and find diverse structures that present the same activity on a specific target, but it can be used as well in de novo design or hit to lead optimization, scaffold hopping and multi-target drug design.²⁶

Moreover, pharmacophore models can be generated either in a structure-based or ligand-based approach.

2.4.2 Structure-based approach

A structure-based pharmacophore model is based on an exploration of the three-dimensional structure of a macromolecule or a complex of the macromolecule with a known ligand. The complementarity of chemical features is investigated, as well as the

possible interactions and steric hindrances (excluded volumes) between receptor and ligand are taken into account.

2.4.3 Ligand-based approach

For a ligand-based model, there is no information about the structure of the target available, but several ligands are known that share a similar biological activity. By generating a model, common feature patterns shared by active ligands are characterized.²⁵

Different software are commercially available to implement pharmacophore models, such as LigandScout²⁶, Phase²⁸, MOE²⁹ or Catalyst²⁹.

2.5 Molecular Docking

2.5.1 General Information

Molecular docking is a computational technique used to predict the binding mode of a ligand to a protein of known three-dimensional structure.³⁰ In general, two methods of docking can be distinguished: protein-protein docking and small-molecule – protein docking, the latter one being more frequently used in pharmaceutical drug-development.

It is acknowledged that there are three main aims of a docking process³¹:

- 1.) The accurate prediction of the active site of the target molecule, the orientation of the ligand and also both their conformations.
- 2.) A ranking of the different compounds according to their binding affinities.
- 3.) To generate a shortlist of potential candidates to test experimentally.

2.5.2 Search strategy

With a variety of search algorithms, it is possible to explore the conformational space of ligand and protein interactions. Different strategies, like systematic search algorithms, molecular dynamics simulations, shape complementarity or genetic algorithms are used to predict the native conformation of a ligand receptor complex.

Due to the high flexibility of the most organic compounds the exploration of conformational space can be challenging, since they can contain many conformational degrees of freedom.³² A good search strategy should find an accurate binding pose without

taking too much time, since thousands of compounds are evaluated during one docking run.

2.5.3 Scoring functions

The goal of scoring functions is to estimate the free energy of binding of a ligand-receptor complex in an aqueous solution.³³

Furthermore, scoring functions should be able to define correct poses from incorrect ones or active molecules from inactives.³⁴ The output of a scoring is a list of ligands ranked according to their estimated binding affinity.

In general, scoring functions can be categorized into force-field based, empirical and knowledge-based functions.³⁴

Force-Field based functions

This method is based on an energy calculation by using a molecular mechanics force field. It includes the sum of non-bonded interactions such as electrostatic and van der Waals potentials as well as contributions from internal distances, angles and torsions.³⁶ However, hydrophobic effects and entropy contributions are not considered in these calculations.

One advantage of this method is the speed, because force fields are able to anticipate many calculation iterations and can estimate the binding energy in a rapid way.³⁵

Empirical functions

In this method, the free energy of binding is decomposed into a function including terms for hydrogen-bonds, hydrophobic interactions, ionic contacts and entropic effects. The different terms are normalized by weighting the coefficients derived from regressions over sets of well-known protein-ligand complexes.³⁵

Knowledge-based functions

As the name implies, knowledge-based methods exploit the knowledge of public databases. Nowadays much information about resolved protein-ligand complexes can be retrieved by using databases such as the PDB. The method is based on a statistical analysis and assumes that the more frequently an interaction occurs, the more favorable it is. Frequent interactions are considered to have a positive contribution, while less frequent contacts are considered repulsive.³⁵ Knowledge based methods strongly

depend on statistical methods, as well as the quality and also the variety of the databases used.³⁵

The last possibility displays the method of consensus scoring, which combines the previously mentioned scoring functions.

Moreover, in a docking the flexibility of both binding partners has to be considered. In general, there exist three forms of docking: rigid docking, semi-flexible docking and flexible docking.

Rigid docking is based on a volume and surface complementarity of ligand and protein.³⁶ It is a simple model which considers protein and ligand as stiff and solid molecules.³⁷ The algorithm calculates the match of shape of both binding partners. The ligand conformations are generally calculated prior to the docking.

In a semi-flexible docking, usually the ligand is regarded flexible, while the protein remains in its rigid conformation.

The third variation, called flexible docking, leaves protein and ligand as flexible molecules. It allows both molecules to change conformations during the docking process, but only to a certain limitation. Mostly, only side chains of the binding site are optimized in their orientation. This approach is based on interaction energy.³⁶

The applicability of docking expands from hit identification, where molecules that fit into the binding cavity are predicted, to lead optimization, where the information of a drug binding to its target is used to design analogues with optimized interactions.

Several programs can be used to conduct a docking, for example GOLD³⁸, FleXX³⁹, DOCK⁴⁰, Autodock⁴¹ and Glide⁴².

2.6 Binding Affinity

The binding affinity describes the strength of the interactions between a ligand and protein.⁴³ Several chemical and physical forces contribute to a binding between a ligand and macromolecule, most of all are electrostatic forces like hydrogen bonds, salt bridges or van der Waals interactions, but also pi-pi interactions and hydrophobic effects can occur.

2.6.1 Thermodynamical Background

When a complex (PL) between ligand (L) and protein (P) is formed, the following equation determines the reaction at equilibrium:



To express the binding affinity of the complex PL, usually the dissociation constant K_d is used to define the stability of the complex, where [P], [L] and [PL] are the equilibrium concentrations of free protein, free ligand, and the complex of both respectively:

$$K_d = \frac{[P][L]}{[PL]}$$

It is important to note here that K_d only works for a reaction at equilibrium. It defines the concentration of the ligand [L] that occupies half of the protein [P] in an equilibrium state. The lower the dissociation constant of the compound, the higher is its affinity and the stronger are interactions between ligand and protein.

Several methods can be used to experimentally measure the binding affinity, e.g. isothermal titration calorimetry, surface plasmon resonance or fluorescence spectroscopy.⁴⁴ K_d can be related to the change in free energy ΔG° using the equation below, which includes also the gas constant R and temperature T:³⁵

$$\Delta G = \Delta G^\circ - RT \ln K_d$$

A computational attempt to estimate the binding affinity is mainly based on the thermodynamic forces that are involved in a binding process.

Based on the Gibbs-Helmholtz equation, the binding process is composed of:

$$\Delta G = \Delta H - T\Delta S$$

ΔGfree energy of binding

ΔHenthalpy

T.....temperature (°K)

ΔSentropy

Binding can be both enthalpy-driven or entropy-driven. Entropic effects arise for example from the ligand that moves into the binding site, thereby releasing highly ordered water molecules into the bulk water. A decrease of the order of a system leads to an increase in entropy. Conversely, enthalpic effects are for instance gained by releasing water molecules from an apolar part of the binding site, where they cannot build hydrogen bonds, to the bulk water where they can form strong interactions.⁴³

For a strong binding affinity, the value of the free energy of binding has to be negative, then the reaction runs spontaneously.

The prediction of binding affinity remains until today very challenging, since many effects are part of this process. A good steric environment (shape of ligand should be complementary to the shape of receptor) has to be considered as well as electronic complementarity, hydrophobic interactions, solvent effects and enthalpy and entropy of the system.

3 Aim of the thesis

Despite the advances in medicine in the past years, the multiple myeloma is still an incurable disease. Especially, high-risk patients are in need of new therapeutic agents to slow down the progress of the disease and prolong their overall survival.

The aim of this thesis was to conduct a structure-based analysis on targets involved in high-risk multiple myeloma, in order to generate a list of potential inhibitors of these targets for a further experimental testing.

The basis of this work was the network model provided by our collaborators around Univ. Prof. Dr. Heinz Ludwig. (see Introduction)

The network model includes 17 target genes identified as high-risk drivers. First, we started to search for X-ray structures of each of the targets in the PDB. For three of them, CREBL2, MYBL2 and ZBTB4 no resolved structures were available. The remaining targets were discussed with our collaborators. Our demands for the computational investigation were a high-resolution of the X-ray structure for precise predictions and also a complexed ligand, to determine the binding site. Our collaborators assisted us with their knowledge in the biological background of the targets and together we limited the number of targets to explore to the following four: MELK, BUB1B, E2F1 and PKMYT1.

Regarding MELK, the maternal embryonic leucine zipper kinase, 29 human protein structures were available on the PDB and indicated that it is an already well studied kinase. For BUB1B, BUB1 Mitotic Checkpoint Serine/Threonine Kinase B, we found four resolved structures, two of them from X-ray experiments and the other two deriving from electron microscopy. For all of them, no complexed inhibitors were available. Therefore, we excluded this target. In case of E2F1, a transcription factor involved in cell cycle regulation, structures were mainly resolved with domains of other proteins. Moreover, we were not able to identify a binding site on structures of E2F1, therefore we had to exclude this target as well.

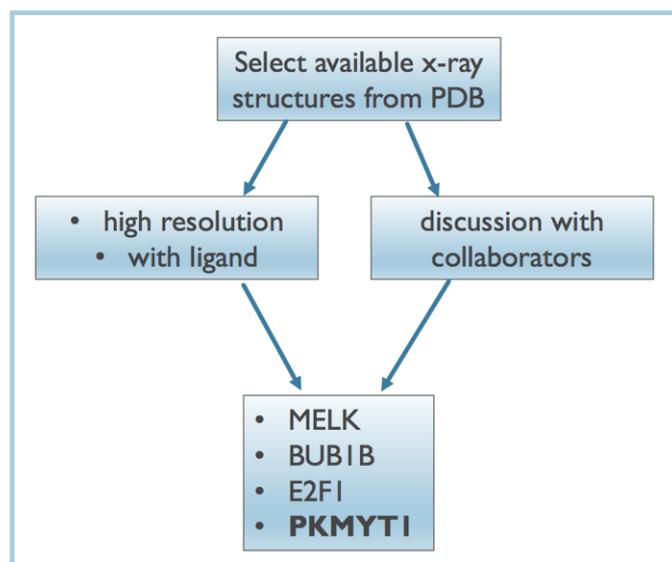


Figure 13: Diagram of our workflow. We started with searching for structures of all targets on the PDB. We wanted a high resolution and also a structure with a complexed ligand. After a discussion with our collaborators we ended up with four targets: MELK, BUB1B, E2F1, PKMYT1. Finally, PKMYT1 was the only target investigated more in detail.

In the end, we decided to focus on only one target, PKMYT1, a member of the WEE family. All members of this family (WEE1, WEE2 and PKMYT1) have an impact on cell cycle regulation therefore we were interested in exploring PKMYT1 more comprehensively.

The purpose of this work was to find potential inhibitors of PKMYT1 using computational techniques. Our main focus of attention was on compounds, which were approved or already tested in clinical trials. This would allow more or less immediate testing in small clinical trials, as safety has already been confirmed.

4 Material and Methods

4.1 Data Collection

We started our structure-based analysis on the targets of the high-risk multiple myeloma pathway by collecting public information about them. For each target, our main questions were: How much information is available for our protein in public databases? Is there a three-dimensional X-ray structure with a high resolution and a complexed ligand of the target protein available? Are there any conserved interactions known? Is there any information about already known inhibitors of the target?

4.1.1 OpenPHACTS

To answer these questions, we searched for the targets in OpenPHACTS, which is a platform that integrates pharmacological data of various different information resources.⁴⁵ The OpenPHACTS analysis delivered information about target pharmacology data, annotated diseases, pathways and also patents for every particular target of the high-risk pathway. The target pharmacology data was extracted entirely from ChEMBL, which is also a free database connected to the OpenPHACTS environment.

4.1.2 Protein Data Bank

The available crystal structures for all seventeen targets were retrieved from the Protein Data Bank (PDB). This task was carried out using a pre-assembled KNIME-workflow, which permits to download the available structures of all targets at once.

4.1.3 ChEMBL

From ChEMBL, a database of bioactive drug-like small molecules, we received information about compounds, which were already tested in biological assays.⁴⁶ ChEMBL provided us K_d and IC₅₀ values for our database of active and inactive molecules.

4.1.4 PubMed

For further general information, we used PubMed, a free website with research literature in life sciences, to get access to important articles about the latest investigations on our targets, especially on PKMYT1.

4.2 Pharmacophore Generation

The pharmacophores were created in LigandScout,²⁵ a computer software of the Inte:Ligand GmbH. Either structure-based or ligand-based pharmacophores can be generated.

We generated a structure-based pharmacophore, based on PKMYT1 bound to dasatinib (PDBID 5VCV⁴⁷). It has been shown that dasatinib presents the strongest binding potential as well as highest selectivity for PKMYT1.^{47,48} Additionally, the X-ray structure 5VCV presents a resolution of 1.92 Å. Thus, this structure is the best candidate to conduct structure-based analysis.

We manually modified the default pharmacophore (Figure 14A) obtained from the PKMYT1-dasatinib complex to make it more generic (Figure 14B). This was required to find potential candidates with distinct scaffolds. Finding the balance between a generic and a specific pharmacophore is challenging, and we tried out several different feature combinations before selecting the most accurate one.

To build this pharmacophore, we first downloaded all available crystal structures in complex with inhibitors (PDB IDs: 5VCV, 5VCW, 5VCX, 5VCY, 5VCZ, 5VD0, 5VD1, and 5VD3) and superimposed them in LigandScout.²⁵ All the inhibitors establish a hydrogen bond with the backbone nitrogen of Cys190 of PKMYT1. Interestingly, literature confirms that the hydrogen bond to Cys190 is conserved.⁴⁸ Thus, this feature was retained in our model. Furthermore, the hydrophobic elements in the back pocket and in the hinge region were also retained, as they were observed for all the inhibitors. Additionally, the two hydrogen bond donors were made optional, as those interactions were present only in a few complexes.

Moreover, recent studies show that a conserved water molecule is building a water bridge with His161 and Glu157.⁴⁹ However, in our structure 5VCV this water molecule was not present and therefore we did not consider it for our model. Another study shows MK1775 building water bridges to Asp251, Tyr121 and Ala137.⁴⁷ In the structure of PDBID 5VCV one oxygen atom of dasatinib can form a water bridge over HOH on position 516 to Asp251, a member of the conserved DFG motif. However, the study does not mention if the interaction with water has an important character for inhibitors, therefore we did not take it into account for our pharmacophore model.

In the original pharmacophore we had 2 hydrogen bond acceptors from two water molecules, but we did not include them in our model, since we did not find enough evidence on their impact on binding from literature and also from our observations in LigandScout²⁵ and Pymol.⁴⁹

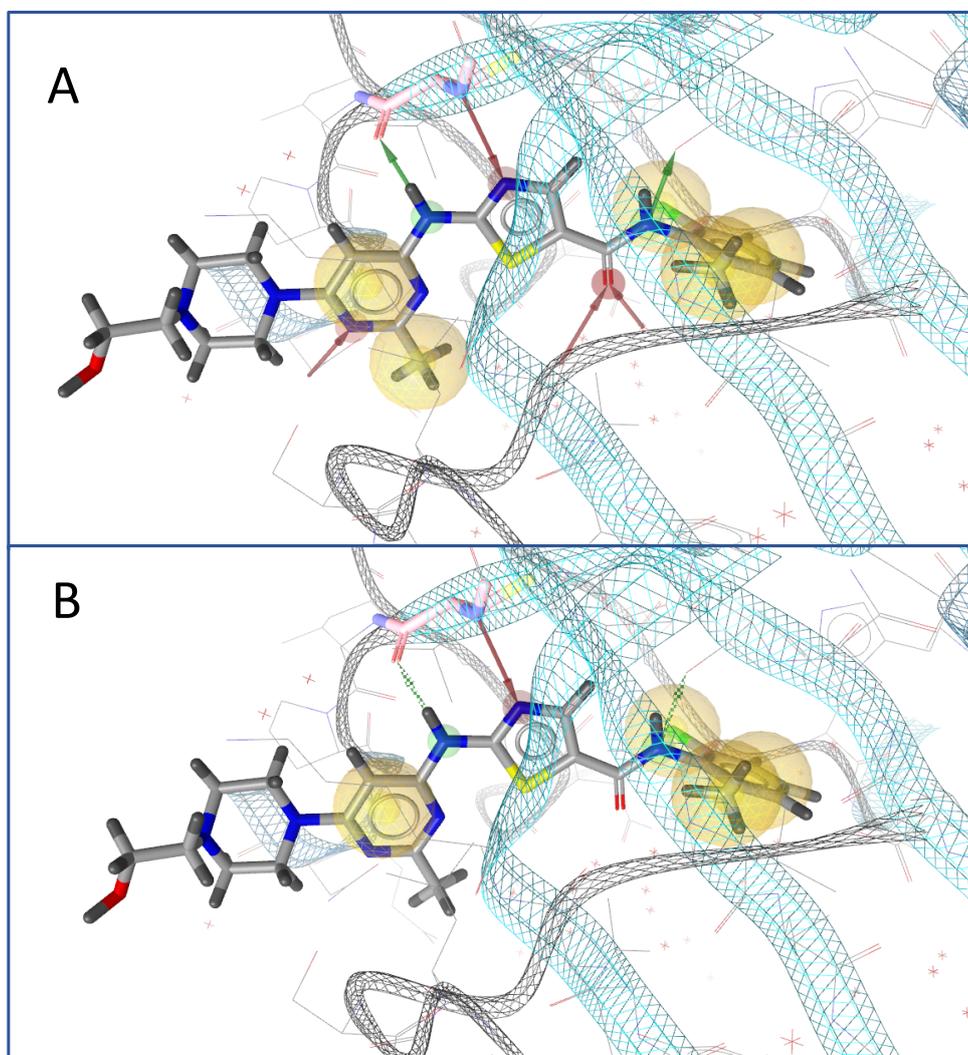


Figure 14: Comparison of pharmacophore models based on dasatinib in complex with PKMYT1 (PDB ID: 5VCV). The default pharmacophore (A), is compared to the generic pharmacophore generated (B). The lipophilic features are displayed by yellow spheres and the hydrogen bond acceptors and donors are depicted by red and green arrows respectively. Green dotted arrows represent optional features. Cys190 is shown as light pink sticks.

4.3 Ligand Docking

The ligand docking was conducted with Glide⁴² from the Schrödinger Suite. Glide⁴² stands for Grid-based Ligand Docking with Energetics. It uses a hierarchical series of

filters which search for the best possible location of a ligand in the active site of the protein and evaluates the ligand's interactions.⁴²

4.3.1 Protein preparation and grid generation

The X-ray structure of PKMYT1 (PDB ID 5VCV) was preprocessed with the Protein Preparation Wizard. This program assigns bond orders and ionization states to the protein prior to the docking.⁵⁰

We used default settings but changed the pH as well to 7.0 ± 0.5 . The binding site was defined by the ligand dasatinib, which is located in the center of the ATP-binding site of the kinase.

The receptor grid was generated with the Receptor Grid Generation tool. This tool takes into account the shape and physico-chemical properties of the binding site. To determine the binding site and generate the grid, we placed the enclosing box in the active site and changed the settings of the ligand midpoint diameter box length in X direction to 12 Å, in Y direction to 21 Å and in Z direction to 10 Å. With these new coordinates of the box, we enhanced the conditions for a docking, by giving the opportunity to find also different binding modes. Here, we set up a hydrogen bond constraint on the backbone of the NH moiety of the Cys190 residue. This constraint was insofar important, as we know that the hydrogen bond interaction to Cys190 is conserved and therefore should be included in all inhibitors.

4.3.2 Ligand preparation

The compounds retrieved from the screening of Drugbank with our generic pharmacophore were imported and prepared with Schrödinger's LigPrep⁵¹ tool. LigPrep converts 1D or 2D structures into low-energy 3D structures and calculates ionization states, tautomers, stereoisomers and ring conformations.⁵¹ Also, it can produce more structures out of the input structure with various ionization states, tautomers, stereoisomers and ring conformations.⁵¹

We used Ionizer to generate protonation states and set the pH to physiological conditions, 7.0 ± 0.5 . The stereoisomer settings were used by default (i.e. the specified chiralities from the source file were retained).

4.3.3 Docking

We performed a flexible docking and recorded one pose per ligand.

During a docking run, Glide searches throughout all generated conformations for the best possible pose of the ligand in the active site of the protein and ranks the ligands

with a scoring function. First of all, the spatial fit of the ligand in the binding site is tested. After this step, the complementarity between ligand and receptor interactions are evaluated. The poses are then minimized with the OPLS-AA force field. Only the final energy-minimized poses get scored by GlideScore, a force-field based scoring function implemented in Schrödinger.⁴²

Due to its operating mode Glide⁴² yields a high accuracy in predicting the binding mode of a ligand, without consuming too much computational costs. The different poses were visually inspected in Pymol.⁴⁹

4.3.4 Redocking PKMYT1

Redocking is a method to validate docking accuracy by reproducing the X-ray binding pose of a ligand in its binding site. First, the PKMYT1-dasatinib complex (PDB ID: 5VCV) was imported and preprocessed with the Protein Preparation Wizard by using default settings, only with the pH changed to 7 ± 0.5 . Then, the ligand dasatinib (ZINC ID: 22005428) was downloaded from the ZINC database and imported into Schrödinger. The ligand preparation included a pH shift to 7 ± 0.5 , as this is the adjustment for a physiological environment. Ionizer was used to generate different ionization states. The receptor grid was created by using the Receptor Grid Generation tool. We used the ATP binding site as active site for our ligand. A flexible docking was performed with default settings, and 5 poses were retained to compare them with the original X-ray binding pose. Moreover, we visually investigated the poses in Pymol.⁴⁹

4.4 Binding affinity calculations

The binding affinity defines how likely a ligand binds to its target protein, thereby considering all chemical and physical effects. With the software SeeSAR, we tried to find out which atoms of the compounds are energetically favorable. SeeSAR is a program for compound prioritization and enhancement.⁵² In fact, not only binding affinity contributions can be determined, it is also possible to explore the whole protein-ligand complex in terms of torsional properties and free space in the binding pocket. Also, it allows the user to change and optimize the structure during a drug development process.⁵²

We imported our protein structure 5VCV and also our docked compounds into the workspace, used the ligand of 5VCV to define the binding site and then calculated the estimated binding energy for each compound.

The program uses its own scoring function called HYDE for these calculations. HYDE is a function based on an estimation of hydrogen bond and dehydration energies that appear during protein-ligand interaction.⁵³ Actually, three contributions are considered: hydrogen bonds, hydrophobic effects and dehydration. For every atom of the ligand-protein interface, the change in energy is calculated.⁵³

It is also very intuitive to use, for example favorable contributions are shown in green spheres, negative contributions are colored in red. Furthermore, the size of the sphere demonstrates the impact of the contribution to the binding – the bigger, the greater is the impact.

We used it just as an additional step to determine our docking poses, find out which structural parts are not favorable and compared it with our other results.

5 Results and Discussion

5.1 PKMYT1

The Protein Kinase Membrane Associated Tyrosine/Threonine 1 (PKMYT1) is a gene, that encodes for a membrane-associated protein kinase, which is located to the ER-Golgi complex.⁵⁴ Kinases are enzymes that phosphorylate their substrates with the help of adenosine triphosphate (ATP). In this process, ATP donates its γ -phosphate group and the kinase catalyzes the transfer of this phosphate to a substrate in its binding site. Actually, only three amino acids - Serine, Threonine, Tyrosine - in the human body can get phosphorylated, as it requires a hydroxy group for this reaction. In general, kinases are divided into two groups: Serine/Threonine kinases and Tyrosine kinases.

PKMYT1 is a negative regulator of the cell cycle and acts as a dual specificity kinase. This type of kinase phosphorylates both Threonines and Tyrosines. Specifically, PKMYT1 phosphorylates Thr14 and Tyr15 of the Cyclin Dependent Kinase 1 (CDK1) and leads thereby to an inhibition of the activity of the CDK1/CyclinB complex, an important participant in the cell's decision to enter mitosis.⁴⁸ If the complex is phosphorylated, it remains in an inactive state and the cell cycle is paused, thereby leading to a G2/M arrest. This arrest allows a DNA-repair, if it is necessary. Afterwards the cycle moves on to mitosis. Normal, non-cancerous cells usually repair defective DNA during a G1 arrest. In cancer cells however, the G1 checkpoint activity is impaired and therefore they deeply rely on the G2/M arrest for DNA repair.⁴⁸ Thus, an inhibition of PKMYT1 can be utilized particularly in cancer cells, to lead them into apoptosis. The best effect regarding anticancer treatment can be reached together with agents that damage DNA⁴⁸. Being part of the high risk multiple myeloma pathway makes PKMYT1 a target worth finding inhibitors for, to reduce the fast progress of the disease.

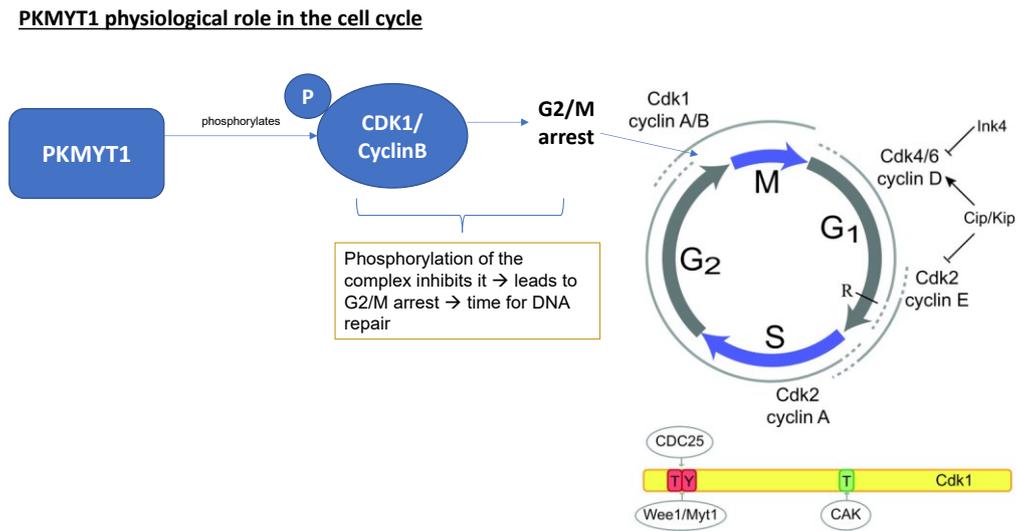


Figure 15: PKMYT1 physiological function in cell cycle. The kinase PKMYT1 phosphorylates its target CDK1/CyclinB, which is inactive in a phosphorylated state. This short arrest however allows a DNA repair and the cell can move on to mitosis. Picture of cell cycle (on the right side) reproduced from Mike Boxem (Cyclin-dependent kinases in *C.elegans*, 2006) and completed with an illustration of the function of PKMYT1.

From the OpenPHACTS analysis (Material and Methods) on PKMYT1, we found annotations to 5 diseases, 8 pathways, 205 counts for pharmacology data and 615 patents in open-source databases.

ID	gene name	patent_count	diseaseCount	pathway_count	targetPharmacologyTotalResults
Q08050	FOXM1	1141	96	4	0
Q99640	PKMYT1	615	5	8	205
P51955	NEK2	1781	17	6	2763
Q01094	E2F1	4659	155	44	11
P10244	MYBL2	1789	25	7	0
Q14680	MELK	1478	13	0	2808
P33981	TTK	2262	23	2	1508
P06493	CDK1	16813	56	39	3612
Q14965	AURKA	6568	141	11	6528
O00444	PLK4	1151	11	6	1343
O60519	CREBL2	222	0	0	0
Q9P1Z0	ZBTB4	235	4	0	0
O43189	PHF1	1150	29	2	0
O60566	BUB1B	1103	41	13	0
Q04206	RELA	3285	192	81	463
Q96HA7	TONSL	168	2	0	0
Q75717	WDHD1	233	1	2	0
P17812	CTPS1	710	1	2	0
Q9BZX2	UCK2	417	12	4	24
P15531	NME1	3517	133	3	11
P04183	TK1	2133	43	4	624

Figure 16: OpenPHACTS results of analysis of genes involved in the high risk multiple myeloma pathway. (Analysis conducted by Jakob Hager)

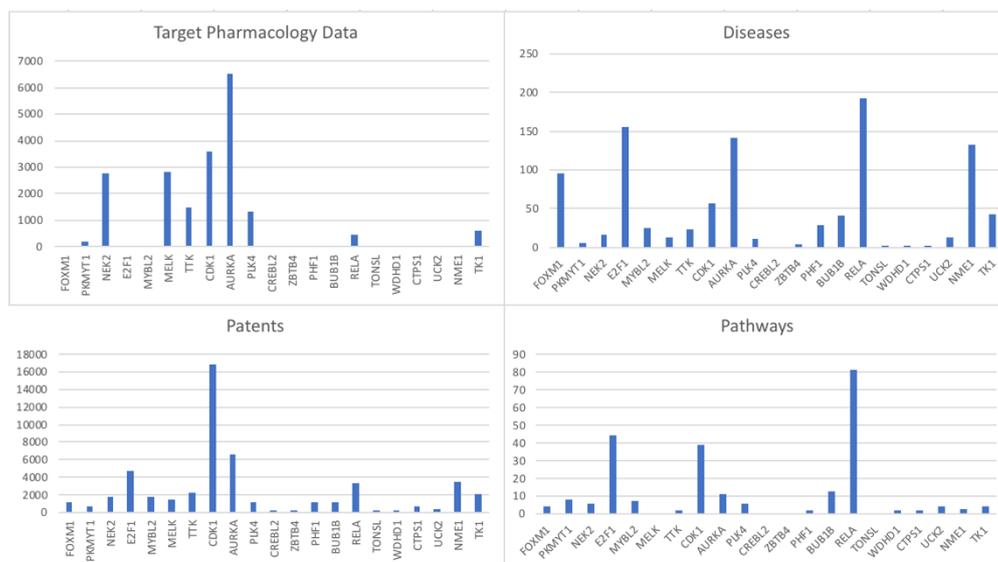


Figure 17: Plots showing data of the OpenPHACTS analysis. (Plots created by Jakob Hager)

In the PDB, nine structures of PKMYT1 were available: one crystal structure in the apo-form and eight crystal structures complexed with inhibitors, i.e. dasatinib, bosutinib, saracatinib, pelitinib, milciclib and MK1775. We compared the complexes regarding their interactions with the protein and came to the following observations:

In all complexes, the inhibitor was establishing a hydrogen bond with the NH of Cys190. From publications we know, that Cys190 is a conserved residue in several kinases of the family of Wee kinases.⁴⁷ Thus, an interaction with this moiety has been considered to be important for inhibition. Furthermore, milciclib, MK1775 and dasatinib, build another hydrogen bond to the oxygen of Cys190.

Additional hydrogen bonds were observed. For example, dasatinib, pelitinib and MK1775 interact with the gatekeeper residue Thr186. Milciclib is the only inhibitor that forms a hydrogen bond with Lys139. Furthermore, dasatinib establishes interaction with Gln196. We concluded that those isolated interactions should not remain in our generic pharmacophore model.

There were also various interactions with water molecules observed. However no specific information from literature was found to include those interactions in our model (for further information see Pharmacophore Generation).

Importantly, all inhibitors possess an aryl ring located in the hydrophobic back-pocket, which demonstrates the importance of hydrophobic interactions in this region.

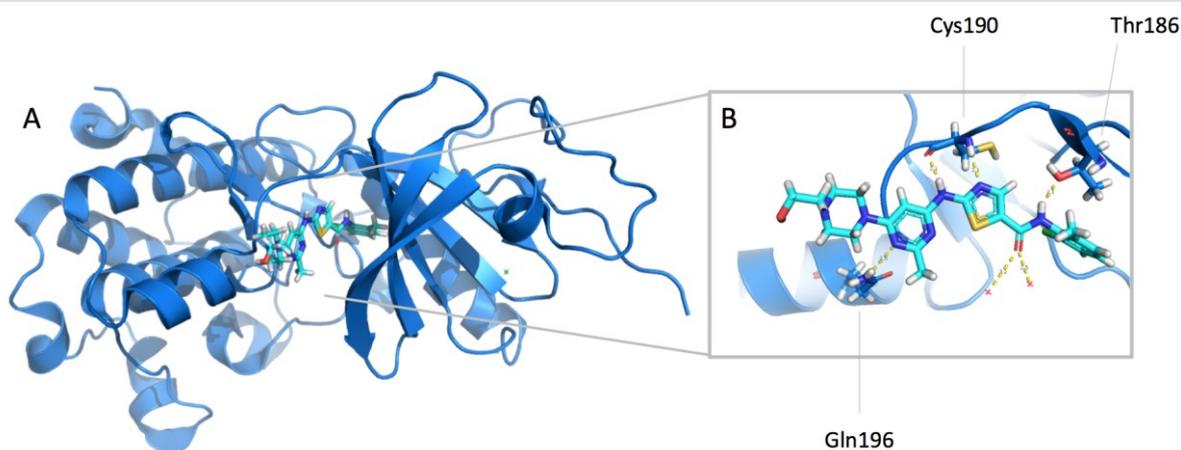


Figure 18: PKMYT1 with its inhibitor dasatinib. Picture A shows an overview of PKMYT1 and in the middle the ATP-binding site is located. In picture B, dasatinib (in turquoise sticks) is shown in the binding pocket. Its interactions (in yellow) with the most important residues (in sticks) and waters (in spheres) are highlighted.

5.2 Pharmacophore Validation, Redocking and Virtual Screening

In a virtual screening, computer programs are used to predict the binding of a compound to a target receptor.⁵⁵ A great number of putative ligands can be screened against a target, with the requirements to have an active site and a protein structure of the target protein available. A virtual screening can be quite fast, depending on the number of ligands being screened, and it does not consume valuable substance material for sophisticated experimental tests, in fact everything is done virtually, with the requirements to have experimental data and computational power.

To find new inhibitors for PKMYT1, we generated first a pharmacophore model based on the most potent inhibitor dasatinib (see Material and Methods). A validation was conducted with the pharmacology data from ChEMBL. We used data of 129 compounds already tested in biological assays. We generated active and inactive databases for the pharmacophore validation using KNIME. The active database was built with compounds with a K_d value below 930 nM and an IC_{50} value below 1200 nM. The inactive database was built with compounds with a K_d value > 10.000 nM. These thresholds were chosen because we had information about binding affinity for active compounds showing values in a nM range. Therefore, we set the threshold for K_d values below 930 nM. For IC_{50} values we inspected the table of binding affinities from

ChEMBL and since values > 10.000 nM indicate a weak or almost no binding potential, we set the threshold at 1200 nM and below for active compounds.

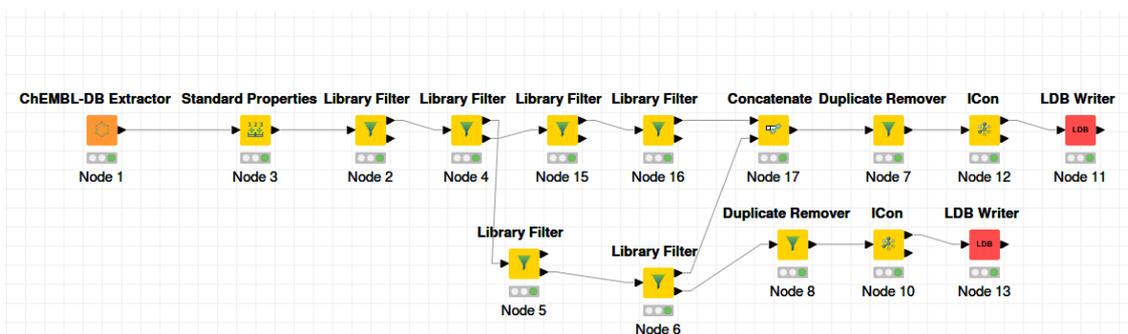


Figure 19: KNIME workflow for the validation. Two separate lists with actives respectively inactives were generated by using bioactivity data of ChEMBL.

We then removed duplicates and generated conformations. The final databases include 6 and 75 compounds in the active and inactive sets respectively. Then, the pharmacophore model was screened against both sets and picked up in total 4 actives and 7 inactives. Several similar pharmacophore models with slight modifications were generated throughout this process without achieving better results.

We validated our model by generating a ROC ('receiver operating characteristics') curve. This is an established metric used to evaluate the ability of a model separating two classes.⁵⁶ In our case, the two classes were active and inactive molecules. Our pharmacophore model should be able to discriminate active from inactive molecules in a database, in order to obtain a high sensitivity and specificity. In the plot, the parameter sensitivity on the y-axis demonstrates the percentage of truly actives divided by the sum of all actives (i.e. true actives and false inactives). The higher the sensitivity is, the better is the model in recognizing all active molecules in a screening.

On the x-axis 1-Specificity is applied, which is the percentage of correctly inactive identified compounds. If specificity is high, the model performs well in discarding inactive compounds.⁵⁶

The ROC curve of the validation for our final model (Figure 20) shows that our model performs better than random, since the AUC is close to the value 1 (0.81). Therefore, our model is satisfyingly able to discriminate between active and inactive molecules.

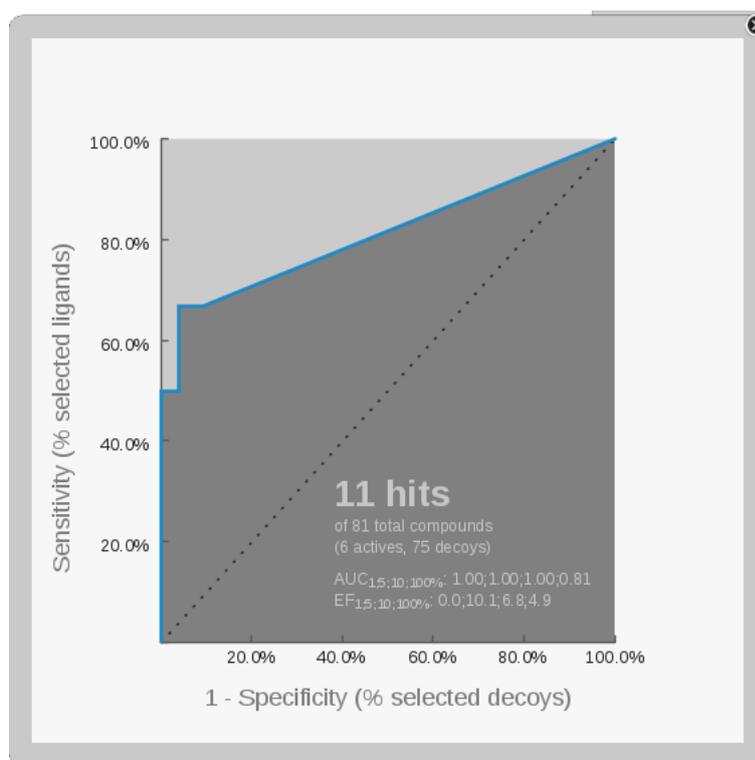


Figure 20: ROC curve of the validation of the pharmacophore model.

Thus, this pharmacophore was used to screen a manually curated dataset including approved, experimental and withdrawn drugs of the DrugBank database. This database is freely accessible and provides detailed drug data and drug target information. For repurposing drugs, the Drugbank is widely used to find already approved drugs which could also be patented for other indications.⁵⁷

From the pharmacophore screening of DrugBank 95 compounds were retrieved. Interestingly, some of these compounds were not kinase inhibitors, e.g. Metipranolol and Bupranolol (both are β -blocking agents), Mebeverine (a gut relaxing agent) and Sitagliptin (DPP4-inhibitor, antidiabetic drug) were also among the results.

We then performed a docking of the 95 compounds using Glide.⁴² To ensure that this program meets the requirements for our research query, a redocking of dasatinib was first conducted (see Material and Methods). Redocking is generally used to validate a docking protocol by checking its ability to reproduce the X-ray binding pose of a ligand in its binding site.

Figure 21 shows the results of the redocking. In fact, it performed very well. In the binding pocket the docked and crystal poses of dasatinib are nearly identical, outside the pocket there are only some slight differences, due to the lack of interactions with

the target kinase. All in all, Glide⁴² fulfilled the expectations for our research question and therefore we used it to execute our docking.

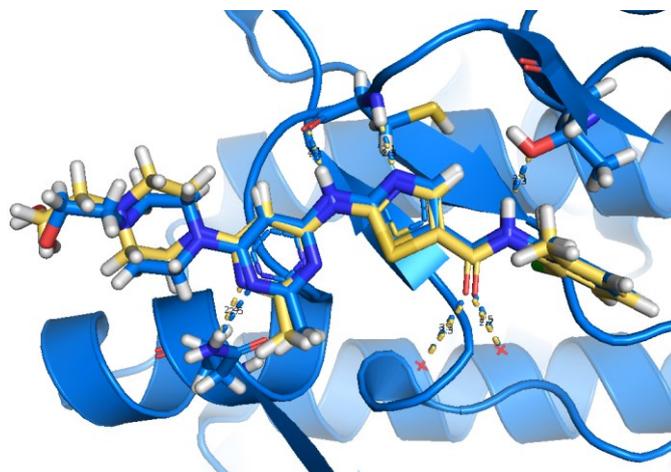


Figure 21: Redocking of dasatinib. The crystallographic complex of PKMYT1 with dasatinib is coloured in blue, and the docking pose of dasatinib is shown in yellow.

Confident in our docking procedure, we then screened the 95 compounds set. From the docking of the 95 compounds resulted a list of in total 106 docking poses. This difference resulted from the LigPrep settings, as we allowed to generate different protonation states between $\text{pH } 7 \pm 0.5$. In addition, we investigated the poses in Pymol. The detailed selection criteria are explained in the next paragraph.

5.3 Compound selection

To select compounds for experimental testing, we visually inspected the binding pose, as well as considered the docking score. Based on our observations of the X-ray complexes of PKMYT1 with inhibitors, we set up a constraint with NH of Cys190.

In our case, the docking score is based on algorithms, which calculate the interactions between the ligand and the protein, how the ligand fits in the active site and the complementarity of ligand-receptor interactions.⁴² We used GlideScore to rank our poses. This score has been developed to distinguish between actives with a high binding affinity from inactives with a low binding affinity.⁵⁸

The redocking of dasatinib presented a score of -11.6. Interestingly, this is the best ranked compound, increasing the confidence in our docking procedure. We primarily

selected 17 compounds based on our analysis of the binding poses and valuable interactions. However, the GlideScore of some of them were quite high. In general, the scores were located in a range between -10.4 and -5.2 for our selected drugs, and although the binding pose and interactions seemed favorable, we were concerned by the proportionally high score.

Therefore, we docked in PKMYT1 the set of 11 compounds obtained from the validation our pharmacophore model. The scores obtained permitted to set a new threshold between real actives and inactives. (Table 1)

Docking of active and inactive compounds

Rank	Name	Activity	Score	
1	Dasatinib	active	-11.660	
2	JNJ-07706621	inactive	-10.600	
3	PD-0173955	active	-9.835	
4	PD-01662 85	active	-9.566	threshold
5	PLX-4720	inactive	-7.884	
6	Brivanib	inactive	-7.497	
7	Brivanib	inactive	-6.742	
8	VX-745	inactive	-6.446	
9	MLN-8054	inactive	-5.351	
10	CI-1040	inactive	-4.690	

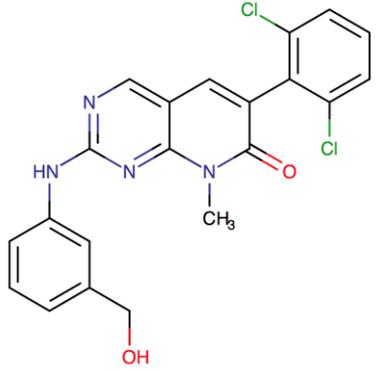
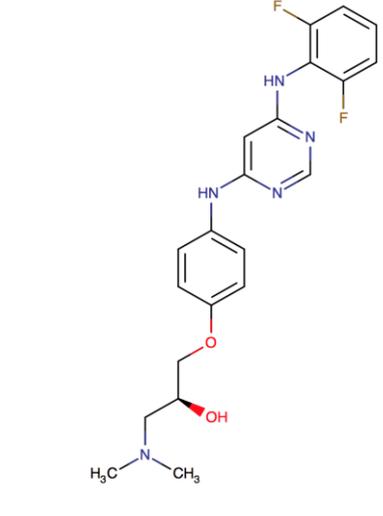
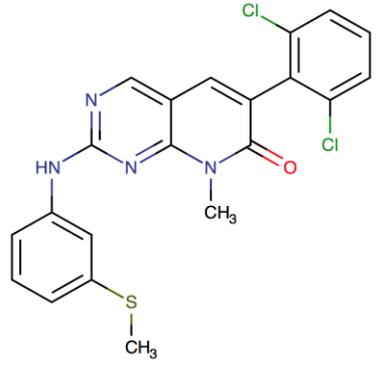
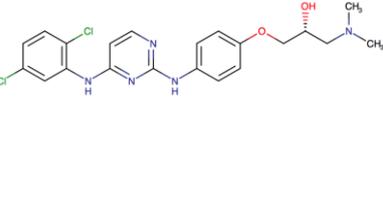
Table 1: Docking of the library used to validate our pharmacophore model. Active molecules are colored in green, inactive molecules in red. It is worth noting, that an inactive molecule is ranked on the second place. We assumed that it is an outlier, since our literature research confirmed that it is an inactive molecule. The pink line depicts our new threshold for the score. One active molecule (Pelitinib) did not appear on this list since it did not fulfill the docking requirements.

It can be observed that the three active compounds have low docking scores. The inactive compound on rank 2 is reported as inactive in literature and for that reason, we stated it as an outlier and did not take it into account for our threshold. In the table can be clearly seen that the new threshold for active compounds lies between -9.5 and -7.8. We updated our list of previously 17 compounds based on this threshold and finally retrieved 8 compounds with scores lower than -7.8. All 8 compounds are experimental drugs and are annotated as targeting a kinase.

As already mentioned, for the last step SeeSAR was used to estimate the binding affinity of our compounds. However, the SeeSAR results were in some terms different to our interpretations from the docking. For example, dasatinib was ranked in SeeSAR on the fifth place, although literature says it is the most potent inhibitor with the highest binding potential among all inhibitors. Also, the docking ranks it on the first place. Furthermore, the conserved interaction to Cys190 is stated as a negative contribution regarding the binding affinity, although every inhibitor from publications shares this interaction. Having this initial situation, we were reluctant to include the results from SeeSAR to our final list. We just want to report that 4 out of the 8 final compounds also appear in top ranks in SeeSAR (highlighted with a * in Table 2). Further conclusions were not taken from the SeeSAR observations.

To sum it up, the following table represents the enhanced version of the best candidates for experimental testing from our docking.

5.3.1 Selected compounds for in vitro testing

Name	Structure	Status	Target	DrugbankID	Glide Score	Rank in Docking
1 6-(2,6-Dichlorophenyl)-2-[[3-(hydroxymethyl)phenyl]amino]phenylamino]-8-methylpyrido[2,3-D]pyrimidin-7(8H)-one		Experimental	Abl1	DB08339	-10,445	2
2 (2S)-1-[4-{{16-[(2,6-Difluorophenyl)amino]-4-pyrimidinyl}amino}phenoxy]-3-(dimethylamino)-2-propanol		Experimental	CDK2	DB07751	-10,014	3
3 6-(2,6-dichlorophenyl)-8-methyl-2-[[3-(methylthio)phenyl]amino]pyrido[2,3-d]pyrimidin-7(8H)-one PD173955		Experimental *	Abl	DB02567	-9,836	4
4 (2R)-1-[4-{{14-[(2,5-Dichlorophenyl)amino]-2-pyrimidinyl}amino}phenoxy]-3-(dimethylamino)-2-propanol		Experimental *	CDK2	DB07750	-9,227	6

Results and Discussion

5	(2R)-1-(dimethylamino)-3-((4-((6-((2-fluoro-5-(trifluoromethyl)phenyl)amino)pyrimidin-4-yl)amino)phenoxy)propan-2-yl)		Experimental	CDK2	DB07054	-9,091	7
6	N-(2,6-dimethylphenyl)-5-phenylimidazo[1,5-a]pyrazin-8-amine		Experimental *	Lck	DB08056	-8,964	9
7	Phenylaminoimidazo[1,2-alpha]pyridine		Experimental *	CDK2	DB04607	-8,815	11
8	[2-Amino-6-(2,6-Difluoro-Benzoyl)-Imidazo[1,2-a]Pyridin-3-yl]-Phenyl-Methanone		Experimental	CDK2	DB04006	-8,672	12

Table 2: Final selected compounds for *in vitro* testing. Targets: Abl= Abelson murine leukemia viral oncogene homolog 1, CDK2=Cyclin dependent kinase 2, Lck= lymphocyte-specific protein tyrosine kinase;

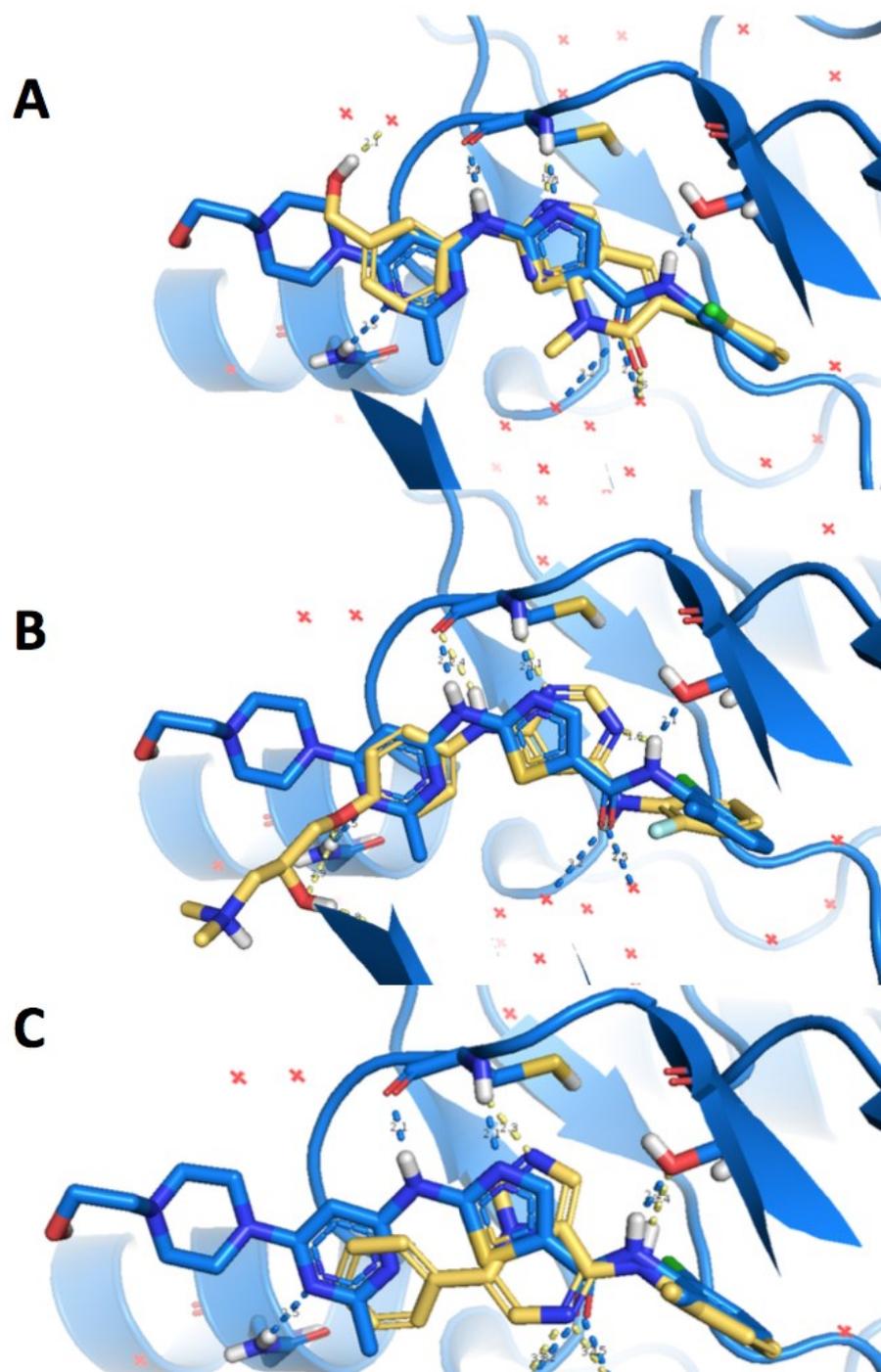


Figure 22: Docking poses of compounds placed on rank 2, 3 and 9. For the compound in picture A (DB08339) with the highest rank in the docking, it can be seen that it fits in the binding pocket and shares the same interactions with Cys190 like dasatinib. Also, a hydrophobic ring is placed in the back pocket. Picture B illustrates the compound DB07751, placed on rank 3, which also shares a hydrogen bond with Cys190 and besides a hydrogen bond with Gln196. In section C, the compound DB08056 is much smaller than dasatinib, but also has several aromatic rings in the same locations, e.g. in the back pocket we have again a hydrophobic part and the molecule shares two hydrogen bond interactions with Cys190 and Thr187.

6 Conclusion

After a computer-aided investigation of the kinase PKMYT1, a member of the high-risk multiple myeloma network model, we generated a list of molecules hypothesized to have an inhibitory effect to this target. Therefore, a pharmacophore model of an already known inhibitor of the target was created and a dataset of experimental, withdrawn and approved drugs of the platform Drugbank was screened with it to find possible inhibiting structures. After a docking of the structures we were able to choose eight reasonable candidates for a further experimental validation. The results from the docking showed well binding poses and also good docking scores for the final candidates compared to the known inhibitor. However, all of the final compounds are categorized as experimental drugs on Drugbank (i.e. they were just tested on animals⁵⁷). The next step would be an in vitro test to verify our predictions.

Since the high-risk multiple myeloma is a fatal disease and better treatment options are urgently needed, research should be continued on the targets of the presented pathway. For example, with our pharmacophore model also various other databases could be screened in order to find new structures of possible inhibitors. Moreover, it would be also beneficial to computationally investigate the other targets of the pathway, as for most of them resolved structures are available in the PDB. In conclusion, it is a very interesting topic worth exploring to eventually obtain better medication for patients suffering from high-risk multiple myeloma with a poor prognosis.

7 Bibliography

1. Greipp, P. R. *et al.* International Staging System for Multiple Myeloma. *JCO* **23**, 3412–3420 (2005).
2. Furchtgott, L. *et al.* Multiple Myeloma Drivers of High Risk and Response to Stem Cell Transplantation Identified By Causal Machine Learning: Out-of-Cohort and Experimental Validation. *Blood* **130**, 3029–3029 (2017).
3. Cooper, G. M. The Development and Causes of Cancer. *The Cell: A Molecular Approach. 2nd edition* (2000).
4. Rajkumar, S. V. & Kumar, S. Multiple Myeloma: Diagnosis and Treatment. *Mayo Clin Proc* **91**, 101–119 (2016).
5. Alant, J., Pool, R., Thomson, J. & Moodley, V. The Diagnosis and Investigation of Multiple Myeloma in 2018. *Archives of Clinical Pathology* **2018**, (2018).
6. IMWG consensus on risk stratification in multiple myeloma | Leukemia. Available at: <https://www-nature-com.uaccess.univie.ac.at/articles/leu2013247>. (Accessed: 6th February 2019)
7. Kumar, S. K. *et al.* Multiple myeloma. *Nature Reviews Disease Primers* **3**, 17046 (2017).
8. Kyle, R. A. & Rajkumar, S. V. Multiple myeloma. *Blood* **111**, 2962 (2008).
9. Plasma Cell Neoplasms (Including Multiple Myeloma). *National Cancer Institute* (1980). Available at: <https://www.cancer.gov/types/myeloma>. (Accessed: 10th January 2019)
10. Mandal, A. & Viswanathan, C. Natural killer cells: In health and disease. *Hematology/Oncology and Stem Cell Therapy* **8**, 47–55 (2015).
11. Nicholson, L. B. The immune system. *Essays Biochem* **60**, 275–301 (2016).
12. Chaplin, D. D. Overview of the Immune Response. *J Allergy Clin Immunol* **125**, S3-23 (2010).
13. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA Cancer J Clin* **68**, 7–30 (2018).
14. Kyle, R. A. *et al.* Review of 1027 patients with newly diagnosed multiple myeloma. *Mayo Clin. Proc.* **78**, 21–33 (2003).
15. Key Statistics for Multiple Myeloma. Available at: <https://www.cancer.org/cancer/multiple-myeloma/about/key-statistics.html>. (Accessed: 12th January 2019)
16. Kumar, S. K. *et al.* Continued improvement in survival in multiple myeloma: changes in early mortality and outcomes in older patients. *Leukemia* **28**, 1122–1128 (2014).
17. International Myeloma Working Group (IMWG) Criteria for the Diagnosis of Multiple Myeloma. *IMWG* (2015). Available at: <http://imwg.myeloma.org/international-myeloma-working-group-imwg-criteria-for-the-diagnosis-of-multiple-myeloma/>. (Accessed: 3rd February 2019)
18. Attal, M. *et al.* A prospective, randomized trial of autologous bone marrow transplantation and chemotherapy in multiple myeloma. Intergroupe Français du Myélome. *N. Engl. J. Med.* **335**, 91–97 (1996).
19. Gertz, M. A. *et al.* Impact of age and serum creatinine value on outcome after autologous blood stem cell transplantation for patients with multiple myeloma. *Bone Marrow Transplant.* **39**, 605–611 (2007).
20. Palumbo, A. *et al.* Autologous transplantation and maintenance therapy in multiple myeloma. *N. Engl. J. Med.* **371**, 895–905 (2014).

21. Durie, B. G. M. *et al.* Bortezomib with lenalidomide and dexamethasone versus lenalidomide and dexamethasone alone in patients with newly diagnosed myeloma without intent for immediate autologous stem-cell transplant (SWOG S0777): a randomised, open-label, phase 3 trial. *Lancet* **389**, 519–527 (2017).
22. Terpos, E. *et al.* European Myeloma Network guidelines for the management of multiple myeloma-related complications. *Haematologica* **100**, 1254–1266 (2015).
23. Lionta, E., Spyrou, G., Vassilatis, D. K. & Cournia, Z. Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. *Curr Top Med Chem* **14**, 1923–1938 (2014).
24. Wermuth, C., Ganellin, C., Lindberg, P., *et al.* (2009). Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure and Applied Chemistry*, 70(5), pp. 1129–1143. Retrieved 17 Sep. 2018, from doi:10.1351/pac199870051129.
25. Wolber, G. & Langer, T. LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. *J. Chem. Inf. Model.* **45**, 160–169 (2005).
26. Pirhadi, S., Shiri, F. & Ghasemi, J. B. Methods and applications of structure based pharmacophores in drug discovery. *Curr Top Med Chem* **13**, 1036–1047 (2013).
27. *Schrödinger Release 2018-4: Phase, Schrödinger, LLC, New York, NY, 2018.*
28. *Molecular Operating Environment (MOE), 2013.08; Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2018.*
29. *Ref. Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 2017, San Diego: Dassault Systèmes, 2016.*
30. Morris, G. M. & Lim-Wilby, M. Molecular docking. *Methods Mol. Biol.* **443**, 365–382 (2008).
31. Bamborough, P. & Cohen, F. E. Modeling protein—ligand complexes. *Current Opinion in Structural Biology* **6**, 236–241 (1996).
32. Kitchen, D. B., Decornez, H., Furr, J. R. & Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews Drug Discovery* **3**, 935 (2004).
33. Hans-Joachim Böhm & Gisbert Schneider. *Virtual Screening for Bioactive Molecules*. **volume 10**, (2000).
34. Meng, X.-Y., Zhang, H.-X., Mezei, M. & Cui, M. Molecular Docking: A powerful approach for structure-based drug discovery. *Curr Comput Aided Drug Des* **7**, 146–157 (2011).
35. Donald J. Abraham & David P. Rotella. *Burger's Medicinal Chemistry, Drug Discovery and Development*. **Volume 2**, (2010).
36. Panda, R., Agrawal, S., Sahoo, M. & Nayak, R. A novel evolutionary rigid body docking algorithm for medical image registration. *Swarm and Evolutionary Computation* **33**, 108–118 (2017).
37. Halperin, I., Ma, B., Wolfson, H. & Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins: Structure, Function, and Bioinformatics* **47**, 409–443 (2002).
38. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking¹¹Edited by F. E. Cohen. *Journal of Molecular Biology* **267**, 727–748 (1997).
39. A fast flexible docking method using an incremental construction algorithm. -

PubMed - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/8780787>. (Accessed: 30th December 2018)

40. Allen, W. J. *et al.* DOCK 6: Impact of new features and current docking performance. *Journal of Computational Chemistry* **36**, 1132–1156 (2015).
41. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J Comput Chem* **30**, 2785–2791 (2009).
42. *Schrödinger Release 2018-3: Glide, Schrödinger, LLC, New York, NY, 2018.*
43. Abraham Donald J. *Burger's Medicinal Chemistry & Drug Discovery. Volume 1*, (2003).
44. Kastritis, P. L. & Bonvin, A. M. J. J. On the binding affinity of macromolecular interactions: daring to ask why proteins interact. *J R Soc Interface* **10**, (2013).
45. About Open PHACTS - Open Pharmacological Space. Available at: <http://www.openphacts.org/about-open-phacts/about-open-phacts>. (Accessed: 26th August 2018)
46. ChEMBL. Available at: <https://www.ebi.ac.uk/chembl/about#>. (Accessed: 22nd August 2018)
47. Zhu, J.-Y. *et al.* Structural Basis of Wee Kinases Functionality and Inactivation by Diverse Small Molecule Inhibitors. (2017). doi:10.1021/acs.jmedchem.7b00996
48. Platzer, C. *et al.* Identification of PKMYT1 inhibitors by screening the GSK published protein kinase inhibitor set I and II. *Bioorganic & Medicinal Chemistry* (2018). doi:10.1016/j.bmc.2018.06.027
49. *The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.*
50. *Schrödinger Release 2018-3: Schrödinger Suite 2018-3 Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2018; Impact, Schrödinger, LLC, New York, NY, 2018; Prime, Schrödinger, LLC, New York, NY, 2018.*
51. *Schrödinger Release 2018-3: LigPrep, Schrödinger, LLC, New York, NY, 2018.*
52. *SeeSAR version 8.1; BioSolveIT GmbH, Sankt Augustin, Germany, 2018, www.biosolveit.de/SeeSAR.*
53. Schneider, N., Lange, G., Hindle, S., Klein, R. & Rarey, M. A consistent description of HYdrogen bond and DEhydration energies in protein–ligand complexes: methods behind the HYDE scoring function. *J Comput Aided Mol Des* **27**, 15–29 (2013).
54. PKMYT1 Gene - GeneCards | PMYT1 Protein | PMYT1 Antibody. Available at: <https://www.genecards.org/cgi-bin/carddisp.pl?gene=PKMYT1>. (Accessed: 19th August 2018)
55. Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery Today* **11**, 580–594 (2006).
56. Triballeau, N., Acher, F., Brabet, I., Pin, J.-P. & Bertrand, H.-O. Virtual Screening Workflow Development Guided by the “Receiver Operating Characteristic” Curve Approach. Application to High-Throughput Docking on Metabotropic Glutamate Receptor Subtype 4. *J. Med. Chem.* **48**, 2534–2547 (2005).
57. Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research* **46**, D1074–D1082 (2018).
58. Docking and Scoring | Schrödinger. Available at: <https://www.schrodinger.com/science-articles/docking-and-scoring>. (Accessed: 22nd August 2018)

8 Appendix

8.1 List of Abbreviations

Abl	Abelson murine leukemia viral oncogene homolog 1
ASCT	autologous stem cell transplantation
CADD	computer aided drug design
CDK	cycline dependent kinase
HTS	high throughput screening
IFM	Intergroup Francophone du Myelome
IGH	immunoglobuline heavy chain
IMWG	International Myeloma Working Group
IUPAC	International Union of Pure and Applied Chemistry
HLA	human leucocyte-associated antigens
Lck	lymphocyte-specific protein tyrosine kinase
MGUS	monoclonal gammopathy of undetermined significance
MM	multiple myeloma
MMRF	Multiple Myeloma Research Foundation
M-protein	monoclonal protein
NCI	National Cancer Institute of America
NK cells	natural killer cells
NSAID	non-steroidal anti-inflammatory drug
OPLS-AA	all-atom optimized potentials for liquid simulations
PDB	Protein Databank
PKMYT1	Protein Kinase Membrane Associated Tyrosine/Threonine 1
ROC	receiver operating characteristics – curve
RANKL	nuclear factor kappa B ligand
SMM	smouldering multiple myeloma

8.2 List of Figures

FIGURE 1: DIFFERENTIATION OF BLOOD STEM CELLS. SOURCE: HTTPS://WWW.CANCER.GOV/IMAGES/CDR/LIVE/CDR596562-750.JPG	- 11 -
FIGURE 2: THE INNATE AND ACQUIRED IMMUNE SYSTEM. SOURCE: HTTPS://WWW.CREATIVE-DIAGNOSTICS.COM/INNATE-AND-ADAPTIVE-IMMUNITY.HTM	- 12 -
FIGURE 3: DIFFERENCE BETWEEN NORMAL PLASMA CELLS AND MALIGNANT PLASMA CELLS (OR MYELOMA CELLS). NORMAL PLASMA CELLS PRODUCE IMMUNOGLOBULINS (=ANTIBODIES). CONTRARILY, THE MULTIPLE MYELOMA CELLS THAT ARE BIGGER IN SIZE AND UNABLE TO PRODUCE EFFECTIVE ANTIBODIES. IN FACT, THEY PRODUCE M-PROTEIN, A SORT OF ABNORMAL IMMUNOGLOBULIN FRAGMENT WHICH LEADS TO IMPAIRED FUNCTION OF THE IMMUNE SYSTEM, BLOOD THICKNESS AND KIDNEY DAMAGE. (PICTURE REPRODUCED FROM: HTTPS://WWW.CANCER.GOV/IMAGES/CDR/LIVE/CDR763079-750.JPG)	- 14 -
FIGURE 4: OVERVIEW OF THE DEVELOPMENT AND PREVIOUS STAGES OF THE MULTIPLE MYELOMA. (PICTURE REPRODUCED FROM KUMAR, S. K. ET AL. MULTIPLE MYELOMA, NATURE REVIEWS DISEASE PRIMERS 3 , 2017).....	- 15 -
FIGURE 5: DIFFERENTIAL DIAGNOSIS CRITERIA FOR MONOCLONAL GAMMOPATHY OF UNDETERMINED SIGNIFICANCE, SMOULDERING MULTIPLE MYELOMA AND MULTIPLE MYELOMA. CRITERIA REPRODUCED FROM RAJKUMAR SV, KUMAR S. MULTIPLE MYELOMA: DIAGNOSIS AND TREATMENT. MAYO CLIN PROC. 2016;91(1):101-19.....	- 16 -
FIGURE 6: THERAPEUTIC SCHEME FOR MULTIPLE MYELOMA, SUBDIVIDED INTO STANDARD RISK, INTERMEDIATE RISK AND HIGH-RISK PATIENTS AND AUTOLOGOUS HEMATOPOIETIC STEM CELL TRANSPLANTATION (ASCT) ELIGIBILITY OR INELIGIBILITY. VRd (= BORTEZOMIB + LENALIDOMIDE + DEXAMETHASONE); KRd (= CARFILZOMIB, LENALIDOMIDE, DEXAMETHASONE); ASCT (= AUTOLOGOUS STEM CELL TRANSPLANTATION) (IMAGE REPRODUCED FROM KUMAR, S. K. ET AL. MULTIPLE MYELOMA, NATURE REVIEWS DISEASE PRIMERS 3 2017)	- 18 -
FIGURE 7: DRUG-CLASSES AND DRUGS USED IN THERAPY OF MULTIPLE MYELOMA.	- 19 -
FIGURE 8: INVESTIGATIONAL DRUGS FOR MULTIPLE MYELOMA.	- 20 -
FIGURE 9: NETWORK MODEL DISPLAYING TARGET GENES INVOLVED IN HIGH-RISK MULTIPLE MYELOMA. PICTURE REPRODUCED FROM FURCHTGOTT, L. ET AL. MULTIPLE MYELOMA DRIVERS OF HIGH RISK AND RESPONSE TO STEM CELL TRANSPLANTATION IDENTIFIED BY CAUSAL MACHINE LEARNING: OUT-OF-COHORT AND EXPERIMENTAL VALIDATION. BLOOD 130 , 3029–3029 (2017)	- 21 -
FIGURE 10: DRUG DISCOVERY PROCESS INVOLVING COMPUTATIONAL TECHNIQUES. PICTURE REPRODUCED FROM GREGORY SLIWOSKI ET AL. (COMPUTATIONAL METHODS IN DRUG DISCOVERY, 2013)	- 23 -
FIGURE 11: OUR WORKFLOW TO FIND NEW INHIBITORS.	- 25 -
FIGURE 12: (A) PHARMACOPHORE MODEL BASED ON KINASE INHIBITOR DASATINIB. GENERATED WITH LIGANDSCOUT. (B) FEATURE EXPLANATIONS FOR LIGANDSCOUT.	- 26 -
FIGURE 13: DIAGRAM OF OUR WORKFLOW. WE STARTED WITH SEARCHING FOR STRUCTURES OF ALL TARGETS ON THE PDB. WE WANTED A HIGH RESOLUTION AND ALSO A STRUCTURE WITH A COMPLEXED LIGAND. AFTER A DISCUSSION WITH OUR COLLABORATORS WE ENDED UP WITH FOUR TARGETS: MELK, BUB1B, E2F1, PKMYT1. FINALLY, PKMYT1 WAS THE ONLY TARGET INVESTIGATED MORE IN DETAIL	- 33 -
FIGURE 14: COMPARISON OF PHARMACOPHORE MODELS BASED ON DASATINIB IN COMPLEX WITH PKMYT1 (PDB ID: 5VCV). THE DEFAULT PHARMACOPHORE (A), IS COMPARED TO THE GENERIC PHARMACOPHORE GENERATED (B). THE LIPOPHILIC FEATURES ARE DISPLAYED BY YELLOW SPHERES AND THE HYDROGEN BOND ACCEPTORS AND DONORS ARE DEPICTED BY RED AND GREEN ARROWS RESPECTIVELY. GREEN DOTTED ARROWS REPRESENT OPTIONAL FEATURES. CYS190 IS SHOWN AS LIGHT PINK STICKS... -	36 -
FIGURE 15: PKMYT1 PHYSIOLOGICAL FUNCTION IN CELL CYCLE. THE KINASE PKMYT1 PHOSPHORYLATES ITS TARGET CDK1/CYCLINB, WHICH IS INACTIVE IN A PHOSPHORYLATED STATE. THIS SHORT ARREST HOWEVER ALLOWS A DNA REPAIR AND THE CELL CAN MOVE ON TO MITOSIS. PICTURE OF CELL CYCLE (ON THE RIGHT SIDE) REPRODUCED FROM MIKE BOXEM (CYCLIN-DEPENDENT KINASES IN C.ELEGANS, 2006) AND COMPLETED WITH AN ILLUSTRATION OF THE FUNCTION OF PKMYT1.	- 41 -
FIGURE 16: OPENPHACTS RESULTS OF ANALYSIS OF GENES INVOLVED IN THE HIGH RISK MULTIPLE MYELOMA PATHWAY. (ANALYSIS CONDUCTED BY JAKOB HAGER)	- 41 -
FIGURE 17: PLOTS SHOWING DATA OF THE OPENPHACTS ANALYSIS. (PLOTS CREATED BY JAKOB HAGER).....	- 42 -
FIGURE 18: PKMYT1 WITH ITS INHIBITOR DASATINIB. PICTURE A SHOWS AN OVERVIEW OF PKMYT1 AND IN THE MIDDLE THE ATP-BINDING SITE IS LOCATED. IN PICTURE B, DASATINIB (IN TURQUOISE STICKS) IS SHOWN IN THE BINDING POCKET. ITS INTERACTIONS (IN YELLOW) WITH THE MOST IMPORTANT RESIDUES (IN STICKS) AND WATERS (IN SPHERES) ARE HIGHLIGHTED. ... -	43 -
FIGURE 19: KNIME WORKFLOW FOR THE VALIDATION. TWO SEPARATE LISTS WITH ACTIVES RESPECTIVELY INACTIVES WERE GENERATED BY USING BIOACTIVITY DATA OF ChEMBL.	- 44 -
FIGURE 20: ROC CURVE OF THE VALIDATION OF THE PHARMACOPHORE MODEL.	- 45 -

- FIGURE 21: REDOCKING OF DASATINIB. THE CRYSTALLOGRAPHIC COMPLEX OF PKMYT1 WITH DASATINIB IS COLOURED IN BLUE, AND THE DOCKING POSE OF DASATINIB IS SHOWN IN YELLOW. - 46 -
- FIGURE 22: DOCKING POSES OF COMPOUNDS PLACED ON RANK 2, 3 AND 9. FOR THE COMPOUND IN PICTURE A (DB08339) WITH THE HIGHEST RANK IN THE DOCKING, IT CAN BE SEEN THAT IT FITS IN THE BINDING POCKET AND SHARES THE SAME INTERACTIONS WITH CYS190 LIKE DASATINIB. ALSO, A HYDROPHOBIC RING IS PLACED IN THE BACK POCKET. PICTURE B ILLUSTRATES THE COMPOUND DB07751, PLACED ON RANK 3, WHICH ALSO SHARES A HYDROGEN BOND WITH CYS190 AND BESIDES A HYDROGEN BOND WITH GLN196. IN SECTION C, THE COMPOUND DB08056 IS MUCH SMALLER THAN DASATINIB, BUT ALSO HAS SEVERAL AROMATIC RINGS IN THE SAME LOCATIONS, E.G. IN THE BACK POCKET WE HAVE AGAIN A HYDROPHOBIC PART AND THE MOLECULE SHARES TWO HYDROGEN BOND INTERACTIONS WITH CYS190 AND THR187. - 51 -

8.3 List of Tables

- TABLE 1: DOCKING OF THE LIBRARY USED TO VALIDATE OUR PHARMACOPHORE MODEL. ACTIVE MOLECULES ARE COLORED IN GREEN, INACTIVE MOLECULES IN RED. IT IS WORTH NOTING, THAT AN INACTIVE MOLECULE IS RANKED ON THE SECOND PLACE. WE ASSUMED THAT IT IS AN OUTLIER, SINCE OUR LITERATURE RESEARCH CONFIRMED THAT IT IS AN INACTIVE MOLECULE. THE PINK LINE DEPICTS OUR NEW THRESHOLD FOR THE SCORE. ONE ACTIVE MOLECULE (PELITINIB) DID NOT APPEAR ON THIS LIST SINCE IT DID NOT FULFILL THE DOCKING REQUIREMENTS. - 47 -
- TABLE 2: FINAL SELECTED COMPOUNDS FOR IN VITRO TESTING. TARGETS: ABL= ABELSON MURINE LEUKEMIA VIRAL ONCOGENE HOMOLOG 1, CDK2=CYCLIN DEPENDENT KINASE 2, LCK= LYMPHOCYTE-SPECIFIC PROTEIN TYROSINE KINASE; - 50 -

Ich habe mich bemüht, sämtliche Inhaber der Bildrechte ausfindig zu machen und ihre Zustimmung zur Verwendung in dieser Arbeit eingeholt. Sollte dennoch eine Urheberrechtsverletzung bekannt werden, ersuche ich um Meldung bei mir.