The current drug therapy possibilities for patients with solid tumors are inadequate, and the drug candidates tend to fail in the clinical trials due to the lack of efficacy. In this thesis three promising drug targets, AurkA, ATX and K-Ras, were evaluated from the perspective of early phase drug design in the respect of their potential efficacy caveats. This thesis provides novel insights to the efficacy problems among these targets and how these problems could be addressed in the early phase drug design, which could lead to lower drug attrition rates related to the lack of efficacy.
Designing Effective Lead Molecules against Solid Tumors with Inadequate Treatment Options
TATU PANTSAR

Designing Effective Lead Molecules against Solid Tumors with Inadequate Treatment Options

The Efficacy Problem

To be presented by permission of the Faculty of Health Sciences, University of Eastern Finland for public examination in auditorium CA102, Canthia building, Kuopio, on Thursday, April 26th 2018, at 12 noon

Publications of the University of Eastern Finland
Dissertations in Health Sciences
Number 456

School of Pharmacy, Faculty of Health Sciences,
University of Eastern Finland
Kuopio
2018
ABSTRACT

Hepatocellular carcinoma and pancreatic ductal adenocarcinoma are solid tumors with a dismal prognosis. In recent years, the death rates for most of the cancers have been decreasing, but unfortunately the mortality associated with liver and pancreatic cancers has been increasing. Both of these tumors can be treated surgically if identified in their early stages. However, they are usually detected in their late stages, and at that stage, there are no effective therapy options available. Although notable effort has been invested in devising novel therapies against these tumors, the therapies against the solid tumors have tended to fail in clinical trials due to the lack of efficacy.

The lack of efficacy can be further divided into failure to treat the disease and poor pharmacodynamics. Of these, the former is dependent on the drug target, which means that the target needs to be valid. The latter is related to the drug’s ability to cause the desirable effect at the validated target. This thesis focused on pharmacodynamic efficacy from the perspective of early phase drug design i.e. to resolve and identify the potential efficacy caveats for treating these currently untreatable solid tumors with respect to three potential drug targets: Aurora kinase A (AurkA), Autotaxin (ATX) and K-Ras.

Indeed, multiple issues regarding the efficacy were noted. All three drug targets exhibit specific efficacy issues, which are represented individually by specific targets. For instance, K-Ras displays discrepancies between the different position 12 mutations that may result in issues with efficacy. Furthermore, some general issues that are uniform among all targets were also identified. First, all of the targets undergo protein–protein interactions that are unaffected by their enzymatic activity, whereas usually the candidate molecules are simply evaluated by this single-metric (e.g. kinase inhibition). A good example of this is the AurkA–Myc interaction, which is not dependent on the AurkA kinase activity. Thus, only those inhibitors that are able to prevent this protein–protein interaction (in addition to the kinase inhibition) display efficacy in a particular subtype of HCC. Therefore, in order to achieve efficacy, it is essential to identify the crucial protein–protein interactions during the drug design process. Consequently, the single-metric approach is incapable of assessing the molecule’s true efficacy. Finally, at present, the dynamics of the target protein and the candidate molecule’s effect on the target protein’s dynamics are poorly identified. Importantly, in order to understand and predict a drug candidate’s efficacy, the drug target needs to be treated as a dynamic entity.

To summarize, the focus in the early-phase drug discovery has been on the single-metric approach, which usually results in a lack of efficacy in clinical trials. Combining in silico methods with the experimental data early on in the drug design process can help to identify and evade these putative efficacy caveats.
Pantsar, Tatu
Huoneonnustisteiden syöpäkasvaimien hoitoon suunniteltujen uusien lääkeaineiden tehokkuusongelma ja sen mahdolliset ratkaisumallit.
Itä-Suomen yliopisto, terveyestieden tiedekunta
Publications of the University of Eastern Finland. Dissertations in Health Sciences 456. 2018. 102 s.
ISSN (print): 1798-5706
ISSN (pdf): 1798-5714
ISSN-L: 1798-5706

TIIVISTELMÄ


Alku vaiheen lääkeainekehityksen yksipuolinen lähestymistapa on yleensä johtanut tehottomuuteen kliinisissä kokeissa. Tietokoneavustetun lääkairnesuunnittelun yhdistäminen kokeellisen datan kanssa aikaisessa vaiheessa lääkeaineen kehitysprosessia voi auttaa tunnistamaan tehottomuuden aiheuttajat ja välttämään turhien molekyylien testauta kalliissä kliinisissä kokeissa.

Luokitus: QV 745, WI 735, WI 810
Yleinen suomalainen asiasanasto: lääkiesuunnittelu; maksasyöpä; haimasyöpä; lääkaineet; lääkehoito; farmakodynamika; tehokkuus; vaikuttavuus
"IF IT BLEEDS—WE CAN KILL IT."

ARNOLD SCHWARZENEGGER (DUTCH) – PREDATOR (1987)
Acknowledgements

The three drug targets covered in this thesis were studied in 2014–2017. In addition to this thesis, all three studied targets resulted in publications, in highly or top ranked journals. The research was conducted in Doctoral Program of Drug Research and were funded by University of Eastern Finland. To note, the research with K-Ras and AurkA are still actively ongoing, and I will continue this research as a PostDoc at University of Tübingen, Germany.

First, I want to express my deepest gratitude to my supervisors Tuomo and Antti. Tuomo, I appreciate your guidance with all the technical and practical issues which I faced, especially in the early phase of my work. I am also grateful that you taught and helped me a lot how to visualize things orderly and clearly. Your hawk-eye, which is able to point out even 1 mm differences, is quite exceptional. Antti, there was hardly a situation where you were lack of new ideas, which helped a lot during this process and also helped my mind to evolve. Moreover, I was happy to get your support and comprehensive feedback for my own research ideas. Also, I appreciate your general guidance in the field of science.

I am thankful for Prof. Anna Linusson and Prof. Olli Pentikäinen for reviewing this thesis. Your valuable comments improved it greatly, and also gave me new perspective in the topics handled in this thesis. I am grateful for Professor Maurizio Botta for agreeing to act as an opponent in the public defence. In addition, I want to thank Ewen MacDonald for reviewing the language of this thesis.

I am thankful for all the people that I have collaborated during these years. Many thanks to Jarmo Laitinen (UEF) and Juha Savinainen (UEF) for the fruitful collaboration, which started with the endocannabinoid related research prior to this thesis work, and continued here with the Autotaxin. I would like to thank Prosanta Singha for conducting all the biological assays for the Autotaxin publication. Prof. Lars Zender’s group from University of Tübingen is acknowledged from the collaboration with AurkA and K-Ras related research. Prof. Ilpo Vattulainen is greatly acknowledged from the collaboration related to MD simulation studies with K-Ras. I am really thankful for Sami Rissanen from Tampere University of Technology with the teamwork in the K-Ras project and also for the introductory guidance to the molecular dynamics simulation universe. CSC – IT center for science is acknowledged for the computational resources that were used in this research.

Also, I want to thank the people of the modeling group and the Pharmaceutical Medicinal Chemistry (PMC) group. Special thanks for Mikko Gynther with all the nice discussions with several scientific and non-scientific topics in the coffee-room.

To be able to focus this work during these years, I am grateful for all the great composers that have provided excellent material for this. Especially, Ramin Djawadi and Hans Zimmer (literally) played a huge role during this process.

I want to also thank all of my friends for the refreshing moments and invaluable memories from these years. Exceptional thanks go to Jouni Jukka and Antti Miikki. Also I would like to thank Peliporukka; the missions behind enemy lines or galactic ambitions really helped me to get my thoughts out of science. I thank my brothers for the refreshing trips for the pursuit of the “great pike” in Luonteri. I am also grateful to my parents, for letting me independently follow and discover my own path in the life.

Last but not least, I want to express my sincere gratitude to Ennska for all your support and love throughout this process.

Tübingen, February 14th, 2018

Tatu Pantsar
Contents

1 INTRODUCTION ................................................................................. 1
  1.1 Solid tumors with inadequate treatment options .................. 1

2 LACK OF EFFICACY HINDERS DRUG DEVELOPMENT FOR SOLID TUMORS ................................................. 2
  2.1 New drug treatments fail in clinical trials .................. 2
  2.2 The cause of the lack of efficacy ............................................ 3
  2.3 Strategy to overcome the lack of efficacy .................. 4

3 METHODS ...................................................................................... 5
  3.1 Methods to obtain atomic coordinates for in silico studies ...... 5
    3.1.1 X-ray crystallography ............................................................. 5
    3.1.2 NMR spectroscopy ................................................................. 7
    3.1.3 Cryo-electron microscopy ....................................................... 8
    3.1.4 Homology modeling ................................................................ 9
    3.1.5 Absent structural information ........................................... 10
  3.2 Molecular docking .................................................................... 10
    3.2.1 Basics .................................................................................... 10
    3.2.2 Analysis ................................................................................ 14
    3.2.3 Limitations ............................................................................ 15
  3.3 Molecular dynamics .................................................................. 16
    3.3.1 Basics .................................................................................... 16
    3.3.2 Timescale and protein movement ........................................ 16
    3.3.3 Analysis ................................................................................ 17
  3.4 Hydration site analysis ................................................................. 20
  3.5 Protein–protein interaction prediction in silico .................. 20
  3.6 Methods to analyze protein–protein interactions in vitro ......... 21
    3.6.1 Isothermal titration calorimetry ........................................... 22
    3.6.2 Surface plasmon resonance .................................................. 22
    3.6.3 Co-immunoprecipitation ....................................................... 22
    3.6.4 Kinase activity assays ........................................................... 23

4 CASE STUDIES OF SELECTED DRUG TARGETS ...................... 25
  4.1 Aurora kinase A ......................................................................... 25
    4.1.1 Aurora kinase A as a drug target ....................................... 25
    4.1.2 Aurora kinase A inhibitors .................................................. 30
    4.1.3 The efficacy problem with Aurora kinase A inhibition ... 36
  4.2 Autotaxin .................................................................................... 39
    4.2.1 Autotaxin as a drug target .................................................. 39
    4.2.2 Autotaxin inhibitors ............................................................... 42
    4.2.3 The efficacy problem with Autotaxin ................................ 44
  4.3 K-Ras ......................................................................................... 47
    4.3.1 K-Ras as a drug target .......................................................... 47
    4.3.2 Targeting K-Ras ................................................................. 52
    4.3.3 The complexity of K-Ras position 12 mutations ............... 54
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADP</td>
<td>Adenosine diphosphate</td>
</tr>
<tr>
<td>ARAF</td>
<td>Serine/threonine-protein kinase A-raf</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>ATX</td>
<td>Autotaxin</td>
</tr>
<tr>
<td>AurkA</td>
<td>Aurora kinase A</td>
</tr>
<tr>
<td>AurkB</td>
<td>Aurora kinase B</td>
</tr>
<tr>
<td>AurkC</td>
<td>Aurora kinase C</td>
</tr>
<tr>
<td>BLA</td>
<td>Biologics license application</td>
</tr>
<tr>
<td>BRAF</td>
<td>Serine/threonine-protein kinase B-raf</td>
</tr>
<tr>
<td>CRAF</td>
<td>RAF proto-oncogene serine/threonine-protein kinase</td>
</tr>
<tr>
<td>Cryo-EM</td>
<td>Cryo-electron microscopy</td>
</tr>
<tr>
<td>EC50</td>
<td>Half maximal effective concentration</td>
</tr>
<tr>
<td>GDP</td>
<td>Guanosine diphosphate</td>
</tr>
<tr>
<td>GTP</td>
<td>Guanosine triphosphate</td>
</tr>
<tr>
<td>HCC</td>
<td>Hepatocellular carcinoma</td>
</tr>
<tr>
<td>H-Ras</td>
<td>GTPase H-Ras</td>
</tr>
<tr>
<td>IC50</td>
<td>Half maximal inhibition concentration</td>
</tr>
<tr>
<td>IFD</td>
<td>Induced fit docking</td>
</tr>
<tr>
<td>ITC</td>
<td>Isothermal titration calorimetry</td>
</tr>
<tr>
<td>Kd</td>
<td>Dissociation constant</td>
</tr>
<tr>
<td>Ki</td>
<td>Inhibitory constant</td>
</tr>
<tr>
<td>K-Ras</td>
<td>GTPase K-Ras</td>
</tr>
<tr>
<td>LPA</td>
<td>Lysophosphatidic acid</td>
</tr>
<tr>
<td>LPC</td>
<td>Lysophosphatidylcholine</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular dynamics</td>
</tr>
<tr>
<td>MSM</td>
<td>Markov state model</td>
</tr>
<tr>
<td>NDA</td>
<td>New drug application</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>PDAC</td>
<td>Pancreatic ductal adenocarcinoma</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein data bank</td>
</tr>
<tr>
<td>PI3K</td>
<td>Phosphatidylinositol 4,5-biphosphate 3-kinase</td>
</tr>
<tr>
<td>RALGDS</td>
<td>Ral guanine nucleotide dissociation stimulator</td>
</tr>
<tr>
<td>RA</td>
<td>Ras-associating domain</td>
</tr>
<tr>
<td>RBD</td>
<td>Ras-binding domain</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root-mean-square deviation</td>
</tr>
<tr>
<td>RMSF</td>
<td>Root-mean-square fluctuation</td>
</tr>
<tr>
<td>SPR</td>
<td>Surface plasmon resonance</td>
</tr>
<tr>
<td>TICA</td>
<td>Time-lagged independent component analysis</td>
</tr>
<tr>
<td>TPX2</td>
<td>Targeting protein for Xklp2</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 SOLID TUMORS WITH INADEQUATE TREATMENT OPTIONS

Hepatocellular carcinoma (HCC) is a lethal disease with a poor prognosis. It is the sixth most common cancer in the world with high mortality; it is the third leading cause of death of all the cancers (Ferlay et al. 2010). In the USA, it is in fifth place, and it has been estimated to be responsible for 28,920 deaths in 2017 (American Cancer Society 2017). At present, HCC is lacking effective treatment options. The most common risk factors for the development of HCC are cirrhosis caused by chronic viral infections (hepatitis B virus and hepatitis C virus) and alcohol, obesity and diabetes (El-Serag 2011). Chronic liver disease or cirrhosis is present in 70–90 % of detected HCC cases (El-Serag and Rudolph 2007). The viral infections may exert a synergistic effect with other risk factors (Forner et al. 2012). There are curative options available for HCC in its early stage, such as surgical resection, liver transplantation and percutaneous ablation (Padhya et al. 2013). However, in the advanced state of the disease, which is the stage when two out of every three patients are diagnosed, the only systematic therapy that has prolonged survival in patients is the drug sorafenib. Unfortunately, the survival time in HCC is increased only by three months with the use of sorafenib (Llovet et al. 2008) and in addition, this drug causes severe side-effects and as many as 20–38 % of the patients have had to discontinue the treatment due to drug toxicity (Dekervel et al. 2013). Sorafenib is a tyrosine protein kinase inhibitor, targeting vascular endothelial growth factor receptors 1, 2 and 3 (VEGFRs) and platelet-derived growth factor receptor β (PDGFR-β) (Wilhelm et al. 2004, Chang et al. 2007). In addition, sorafenib inhibits the serine-threonine kinases, CRAF and BRAF.

Pancreatic ductal adenocarcinoma (PDAC), with its 5-year survival rate of a mere 6%, is the most lethal common cancer (Siegel et al. 2015, Yin et al. 2016). In 2016, in the USA alone, PDAC is estimated to be the third most common of all cancer types responsible for mortality (43,090 in a year), only surpassed by the more common lung/bronchus and colon cancers (American Cancer Society 2017). Similar to HCC, PDAC, is lacking effective treatment options. In addition to hereditary syndromes, identified risk factors for PDAC are diabetes mellitus, pancreatitis, smoking, obesity and inactivity (Ryan et al. 2014). These risk factors, excluding hereditary syndromes, increase the risk for PDAC by 2–6 fold. The only potentially curative option for PDAC is surgical resection. This option, however, is suitable only for 15–20% of all patients, as most of the tumors are already locally advanced or metastatic at the time of diagnosis. In advanced or metastatic disease, chemotherapy only improves the survival of patients by a few months.

The pancreatic and the liver cancers are exceptional cancers as their mortality rates are increasing, whereas for other cancers with significant mortality rates, the rates have been declining (Siegel et al. 2015). As the current drug treatment possibilities for HCC and PDAC are inadequate at best, non-existent at worst, these patients are in need of novel therapeutic options. Therefore, there is an urgent demand for innovative approaches to obtain novel lead molecules against HCC and PDAC.
2 Lack of Efficacy Hinders Drug Development for Solid Tumors

2.1 NEW DRUG TREATMENTS FAIL IN CLINICAL TRIALS

Despite the high-demand for novel therapy options for HCC and PDAC, no major breakthroughs have occurred for these diseases. Overall, the drug discovery and development are a challenging process and the success rates are low. If one considers all indications, for drug candidates that end up in the clinical trials (Phase I), the probability for the approval, i.e. that the drug reaches the market, is 9.6% (Thomas et al. 2016). This probability is even lower for cancer drugs, as from all therapeutic areas, oncology drugs display the lowest likelihood (5.1%) of approval from Phase I studies (Fig. 1a). The transition success rates in clinical trials with oncology indications are: Phase I to Phase II 62.8%; Phase II to Phase III 24.6%; Phase III to New drug application (NDA) / Biologics license application (BLA) 40.1%; NDA/BLA to approval 82.4%. From these, the first and the last transition success rates are similar as encountered with non-oncology indications (Fig. 1a). However, oncology has clearly lower rates in advancing from Phase II to Phase III (24.6% vs. 34.3% non-oncology) and from Phase III to NDA/BLA (40.1% vs. 63.7% non-oncology) (Fig. 1a–b). With solid tumors, the phase III success rate is even worse, only 34.2%. Moreover, the worst success rate (13.3%) is observed with pancreatic cancer.

![Graph showing drug candidate success in clinical trials](image)

*Figure 1. Drug candidate success in clinical trials. The drug candidates’ success rate is diminished for solid tumors in phase II and phase III clinical trials (a). Success rates in phase III by selected indications (b). Oncology indications have a clearly diminished success rate, and it is even worse with solid tumors, especially with pancreatic cancer.*

In the 1990’s, issues related to pharmacokinetics were the main reason for drug attrition (Kola and Landis 2004). Nowadays, the main reason for drugs to fail in clinical trials is a lack of efficacy. In Phase II, efficacy is the main reason for attrition, followed by clinical safety and toxicity issues (Waring et al. 2015). The majority of reasons for drug failure (54%) in Phase III are also due to the lack of efficacy (Hay et al. 2014). Overall, it seems that the lack of efficacy is the key-issue with clinical candidates targeting solid tumors.
2.2 THE CAUSE OF THE LACK OF EFFICACY

The lack of efficacy, especially with solid tumors, is truly hindering drug development. In general, the efficacy can be divided into two main concepts: 1) the pharmacodynamic efficacy, and 2) the disease efficacy (Taylor 2014). First, the pharmacodynamic efficacy, is defined by the drug molecule’s ability to exert an effect on its target in vivo. Whereas the disease efficacy is a larger scale efficacy metric, as it is the compound’s ability to affect to the disease. With solid tumors, the disease efficacy may be for instance, tumor regression or an increase in the overall survival.

In terms of disease efficacy, the most crucial factor is the validity of the drug target. If the target is not valid, obviously, disease efficacy cannot be achieved. Nowadays, suitable methods to assess the target validity exist. For example, RNA-interference may be utilized to identify valid drug targets (Zender et al. 2006, Zender et al. 2008). In short, the method blocks one specific protein’s production; thereby, simulating a drug-therapy against the target which results in the target protein’s total inactivity. Moreover, the strength of this method is that one can determine if the function of the target-protein is really crucial for the cancer, or can it be overcome via some other pathway. It is noteworthy, there exist several pitfalls in target validation, which need to be carefully considered (Kaelin 2017). For instance, the association of a protein or its increased level with poor prognosis in a particular cancer does not mean that it is not a good target. Moreover, caution should be put to the used animal models, particularly to the cell-line derived xenograft models, as they appear to translate the efficacy poorly to the clinics (Ocana et al. 2011, Day et al. 2015).

One crucial aspect of the target validity is the druggability of the target (Owens 2007). The term “druggability” refers to the possibility to modulate the target protein with a small-molecule drug. In other words, an undruggable protein lacks potential binding sites in which a small-molecule can bind, thereby meaning that a small-molecule intervention with the protein would be extremely difficult if not impossible. In order to assess druggability, there are tools available to search for potential binding sites from proteins, for example the SiteMap from Schrödinger (Halgren 2009).

Even with a properly validated target, lack of efficacy may be an issue. This appears, if high-enough pharmacodynamic efficacy cannot be achieved to the target. Most often, the small-molecule is evaluated via its binding affinity value (Kd), its EC_{50} or IC_{50} value in a selected assay, as well as some aspects related to selectivity, toxicity and pharmacokinetics. These are indeed crucial properties for a small-molecule. However, the desired action of the molecule is not simply to possess these properties, but in addition to cause the desired effect on the drug target itself. Surprisingly, little attention has been paid to the overall effect of small-molecules on their target. In other words, for example with a kinase protein target, only the inhibition of the kinase activity is usually evaluated. At present, insufficient attention is being paid to the effect of small-molecules on the other plausible protein–protein interactions with the target kinase, unrelated to its kinase activity. Usually, these are only taken into consideration if the drug is acting allosterically i.e. binding to an alternative site.

As noted above, the pharmacodynamic efficacy is usually only evaluated via a single metric, which does not necessarily reflect all of the blocking functionality of the protein, this entirety was observed in the original screen where the protein was identified as a potential target. Overall, there exists a clear gap in the current assessment of pharmacodynamic efficacy.
2.3 STRATEGY TO OVERCOME THE LACK OF EFFICACY

The lack of efficacy, especially with solid tumors, is truly hindering drug development. To overcome the lack of efficacy, one needs to obtain high-enough pharmacodynamic efficacy in order to achieve disease efficacy.

First, as stated above, the drug target needs to be convincingly validated. If that is not conducted, its exploitation in drug discovery will be totally useless, as the disease efficacy cannot be assessed with an invalid target. Moreover, the novel targets usually are not traditional drug targets, such as enzymes or G protein coupled receptors (GPCRs) (Santos et al. 2017). This expansion of druggable targets, however, poses unique challenges for drug discovery programs, as these targets usually lack any obvious small-molecule binding sites and display a plethora of transient protein–protein interactions (Makley & Gestwicki 2013). In summary, resolving the protein–protein interaction challenge (Scott et al. 2016).

Next, as highlighted previously, one needs to identify a compound with a high-enough pharmacodynamic efficacy to the validated target. A potential way to address the lack of efficacy is a thorough characterization of the candidate drug molecule’s effect on its target protein. The hypothesis behind this approach is that by utilizing various *in silico*, *in vitro* and *in vivo* methods, the efficacy of a candidate molecule can be estimated. In other words, one cannot trust on a single metric in the pharmacodynamic evaluation. As most of the potential drug targets display a plethora of interactions with other proteins, the important interactions relating to efficacy need to be identified and the drug candidate’s effect on these processes should be evaluated. Moreover, to aid this evaluation, the publicly available data is constantly expanding and becoming more easily accessible; for instance, UniProt is a hub for protein information and Pharos collates information of protein targets and their ligands (Bateman et al. 2015, Nguyen et al. 2017). Obviously, it is obligatory to ensure that the drug candidate possesses sufficient selectivity as well as a good toxicity and pharmacokinetic profile.

The applied *in vivo* animal model should be also carefully selected. For instance, the most used cell line-derived mouse models do not reflect well with the true characteristics of the cancer (Gengenbacher et al. 2017), and these models further fail to predict the efficacy (Day et al. 2015).

In conclusion, prior to embarking on expensive clinical trials, the drug target, the drug candidate and its action on the target should be evaluated thoroughly. This thesis will cover case studies of three totally different potential drug targets that were studied in 2014–2017. In this context, the issues and the potential issues relating to the efficacy with each drug target will be examined.
3 Methods

3.1 METHODS TO OBTAIN ATOMIC COORDINATES FOR IN SILICO STUDIES

If one wishes to make a direct visualization of a protein of interest at the atomic resolution, one needs to provide the atomic coordinates for the molecular modeling software. These coordinates can be resolved experimentally, or they can be estimated computationally using homology modeling utilizing already existing related structural data. The Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) holds a public archive for resolved biological macromolecular structures and the structural data can be freely downloaded from its website (Berman et al. 2000). Currently, the database consists of nearly 130,000 structures. The majority of the deposited structures in the database have been solved by X-ray crystallography (89%), followed by NMR (9%) with 1% of the structures based on electron microscopy.

The next sections will briefly overview these methods which can be utilized to resolve the atomic coordinates of the protein of interest. Moreover, these methods can be either utilized individually or, especially, with the large macromolecular complexes, the complementary data obtained with each method, means that advantages can be maximized and caveats minimized, and furthermore, they can be combined with computational methods to obtain refined structural information (Goh et al. 2016).

3.1.1 X-ray crystallography

Since the structure of myoglobin was obtained by X-ray crystallography nearly sixty years ago, the technique has become the most common way to study the atomic structure of macromolecules (Kendrew et al. 1958). In the experimental setting, an X-ray beam is targeted at the sample crystal (Berg et al. 2002). This will result in a diffraction pattern as the electrons in the sample scatter X-rays. Furthermore, these scattered waves interfere with each other, either by reinforcing (in phase) or cancelling out (out of phase) each other. The resulting pattern is detected in the X-ray detector. This measurement is repeated numerous times as the crystal is rotated. This results in information about specific intensities and their positions. Based on this obtained diffraction pattern, if the phases are known, the electron density map can be solved via a mathematical conversion, the three-dimensional Fourier transform. Interpretation of this electron density map leads to the atomic coordinates (Fig. 2). Importantly, the phases need to be solved before one can resolve the structure, and this is a central problem in crystallography (Sheldrick 1990, Harrison 1993). The number of structures solved by X-ray in PDB has been steadily increasing, reaching the highest number in the last year: 10,136 structures. In fact, it is the most commonly used method in drug discovery today.
The quality (accuracy) of a crystal structure is described by resolution. The resolution describes the average positional uncertainty of all atoms in the structure. To note, hydrogen atoms scatter X-rays weakly and, therefore, are difficult to be located by the method. Nevertheless, some high-resolution structures (< 1.2 Å) may contain location of some hydrogen atoms that have been resolved. The obtained resolution of a protein structure by X-ray crystallography falls usually in between 1.5 Å and 3.0 Å (Fig. 3). Currently, the highest resolution owing structure in PDB is a small protein crambin structure (PDB ID: 3NIR) with its 0.48 Å resolution (Schmidt et al. 2011).
packing forces may cause artificial stabilization or conformations for local regions in the protein (Rapp and Pollack 2005).

The analysis of the model quality is also an important aspect when evaluating structural data obtained by X-ray crystallography. For instance, free R value (Brünger 1992, Kleywegt and Brünger 1996), that measures the agreement of observed and predicted data is an important parameter in the model quality assessment. Nowadays, selected validation parameters are readily shown in the PDB (Read et al. 2011).

One limitation of the X-ray crystallography method is that the investigated protein structures will need to be able to form crystals. First, this crystallization procedure is not always trivial and has meant that many proteins, especially membrane proteins and protein complexes, cannot be examined with this method. Furthermore, the crystallization conditions may lead to local arbitrary protein conformations. Finally, the obtained crystal conformation is always a spatiotemporal average of the structure.

3.1.2 NMR spectroscopy

Another option to obtain protein coordinates is nuclear magnetic resonance (NMR) spectroscopy. The methodology was pioneered for protein structure analysis by Kurt Wüthrich, resulting in the award of Nobel Prize in Chemistry (Wüthrich 1986, The Nobel Prize in Chemistry 2002). The NMR spectroscopy is based on the fact that specific atomic nuclei are intrinsically magnetic; for instance, $^1$H, $^{13}$C and $^{15}$N (Berg et al. 2002). In the experimental setting, a magnetic field is applied to the sample, resulting in two spin states, $\alpha$ and $\beta$, of the magnetic nuclei. There exists an energy difference between these two states, which is correlated to the strength of the magnetic field. The $\beta$ spin state is an excited high energy state. By applying electromagnetic radiation to the sample, the $\alpha$ spin state will change to the excited $\beta$ spin state. This state transition generates the NMR spectra. Moreover, the surrounding electron density of nuclei may generate a small local magnetic field, which shields a nucleus. This shielding-effect leads to changes in the spin state which are dependent on the chemical surroundings; this means that a nucleus in a different chemical environment will change its spin state with a different radiation frequency or magnetic field strength. In addition, the magnetization is transferred from excited nuclei to unexcited nuclei within a range of 5 Å. This effect is called the nuclear Overhauser effect. With this information, graphical information is obtained of the nuclei which are in close proximity to each other. Based on this proximity data, the three-dimensional structure of the protein can be constructed (Fig. 4). From this data, not just one, but a family of protein structures are generated. This is due to the inadequacy in the experimental data and the dynamic nature of proteins in solution. This may be useful, as it also describes the protein dynamics. The flexible parts of the proteins acquire different conformations within the family, and the more rigid parts of the protein appear to be more stable.
The main advantage of the NMR spectroscopy over X-ray crystallography is that the structure can be in solution i.e. the protein does not need to be crystallized. Traditionally, NMR spectroscopy was limited to small-sized proteins only but today the methodology has advanced to handle larger proteins (Huang & Kalodimos 2017). Moreover, membrane protein structures can be solved via magic angle spinning solid-state NMR spectroscopy (Wang et al. 2013). Furthermore, computational methods can be utilized to build NMR-data driven models. For example, K-Ras lipid-bilayer nanodisc NMR-data driven models investigating K-Ras’ effector-binding site occlusion by the membrane has been published, PDB IDs: 2MSC, 2MSD, 2MSE (Mazhab-Jafari 2015). The number of structures solved by NMR and deposited in PDB has been slowly declining since the year 2007 (RCSB). In 2016, 456 structures solved by NMR were deposited to PDB. The main limitation of the NMR spectroscopy is that the proteins need to be labeled with the specific isotopes, and this may be expensive or not even possible.

3.1.3 Cryo-electron microscopy
Cryo-electron microscopy (cryo-EM) is the latest method applied to resolve protein structures on the atomic scale. In recent years, due to technological and methodological improvements, cryo-EM has been able to produce near-atomic-resolution structures with an increasing frequency, and was selected as the method of the year 2015 by Nature Methods (Nature Methods 2016). Moreover, Nobel Prize in Chemistry was awarded to the developers of the methodology in 2017 (The Nobel Prize in Chemistry 2017).

In the experimental setting, the purified protein (in solution) is applied to a sample grid and instantaneously frozen (Doerr 2016). Next, this frozen protein sample is targeted with an electron beam, and the scattered electrons are detected on the electron detector. From the projections, the particles are aligned and averaged to generate a 3D-map; in other words, the
method depends on computationally averaging of thousands of images of identical particles (Cheng et al. 2015). Importantly, the applied electron radiation damages the organic material but, conversely, with too low radiation one will not obtain a feasible signal-to-noise ratio. Consequently, a movie can be recorded of the experiment, where during the first frames, the electron radiation is kept low, and the amount of radiation is gradually increased.

Traditionally, the near-atomic resolution was achieved with cryo-EM only with structures with molecular weights over 500 kDa and high extensive symmetry (Merino & Raunser 2017). However, the introduction of better detectors, the direct electron detectors, has revolutionized the cryo-EM field. Since 2014, the near-atomic resolution structures, with the resolution below 4 Å, started to emerge. To date, resolution structure below 2 Å was achieved by cryo-EM, e.g. glutamate dehydrogenase, solved with 1.8 Å resolution (PDB ID: 5K12) with the smallest protein (93 kDa) being isocitrate dehydrogenase at 3.8 Å resolution (PDB ID: 5K10) (Merk et al. 2016). Moreover, from the perspective of drug discovery, structural data about important drug targets have been achieved with the method; for instance, the human 80S ribosome was resolved with 3.6 Å resolution (Khatter et al. 2015).

In summary, the methodology is still somewhat limited by the size of the macromolecule, typically a size less than 200 kDa causes challenge for cryo-EM (Merino and Raunser 2017). The number of structures solved by electron microscopy deposited in PDB is constantly increasing, reaching 408 deposited structures in 2016 alone.

### 3.1.4 Homology modeling

If the structural data is unavailable and not feasibly obtainable, one may utilize homology modeling in specific and well considered cases. In homology modeling, an existing template structure with a sufficiently high similarity is needed. The sequence identity with the template and the target protein should be at a minimum of 30% (Martí-Renom et al. 2000). First, related template structures are identified and a feasible template, if available, is selected. For instance, the basic local alignment search tool (BLAST) can be utilized in order to search suitable template structures (Altschul et al. 1990, Boratyn et al. 2013). Next, the sequence of the target protein is aligned with the template sequence(s). This alignment procedure is the crucial step in building the model; thus, it needs to be accurate. Usually, multiple sequences are aligned simultaneously, in order to enhance the alignment, especially of the conserved regions (Edgar & Batzoglou 2006). There are various approaches/software which can be used to conduct the alignment. For example, PROMALS3D utilizes consensus alignment, which is based on both the sequence and the structure (Pei et al. 2008). After the alignment, a model is built, which is subsequently evaluated. The model is created sequentially, first the conserved backbone regions are built, next the loop-regions are added and finally the side-chains are added and the whole model is optimized. Obviously, the resulting model's quality needs to be assessed. There are several ways to evaluate the constructed model. In SWISS-MODEL (Biasini et al. 2014) there is a scoring function QMEAN (Benkert et al. 2011), which evaluates distance-dependent interaction potentials of mean force based on C-β atoms and the long-range interaction on all atom types, torsion angles for local backbone geometry and residue burial, solvent accessibility and the secondary structure prediction. Moreover, there exist several software and web-based sites that can be useful in the evaluation of the structure; for example, ProSA-web (Wiederstein & Sippl 2007). If the model appears invalid, the alignment can be recreated or the template structure can be changed.

Various molecular modeling software include homology modeling tools; namely, MODELLER in Discovery Studio (Dassault Systèmes 2016) and Prime in Schrödinger Maestro (Jacobson et al. 2002, Jacobson et al. 2004). In addition, there are freely available automated servers with web interfaces, such as SWISS-MODEL (Biasini et al. 2014) and I-TASSER (Yang et al. 2015), exist.

The major limitation of homology modeling is that one needs to have similar structure(s) available. Currently, reasonable methods to model a protein without a suitable template structure do not exist. Although, there are software available, such as Rosetta (Kaufmann et al. 2010) and
QUARK (Xu and Zhang 2012), which try to address this by constructing a protein by using a fragment-based building approach.

### 3.1.5 Absent structural information

Quite often a resolved structure, for example solved with the X-ray crystallography, is missing parts of its structural information i.e. it may be lacking some loops from its structure, as it may not have been possible to solve their position with the method, due to the high mobility of these regions. Therefore, if these regions are crucial and needed for the studied system in hand, these regions can be further modelled with suitable software (Goh et al. 2016). For example, Loopy is a tool in JACKAL software which one can utilize in predicting missing loop conformations (Xiang et al. 2002), and in the Schrödinger molecular modeling suite, the tool for modeling the missing loop structures is Prime (Jacobson et al. 2002, Jacobson et al. 2004).

If the missing structural data is too extensive, i.e. if more than just loop region(s) are missing, these missing elements cannot be modeled reliably. Myc is a good example of this type of a protein, only four distinct short regions have been crystallized in a complex with another protein, but most of the structure is still totally unknown (Bayliss et al. 2017). In fact, some proteins appear totally disordered, especially as homomers, and do not seem to have any specific conformation(s) (Marsh & Teichmann 2015). Thereby their structures are unsolvable with the available methods and remain totally unknown to date.

To conclude, as these absent structural information parts of the protein have not been stabilized in a particular conformation, it is evident that modeling these parts in a particular fixed conformation is highly questionable. Although, for example a ligand binding to a specific site adjacent to a flexible region may in some cases stabilize the structure and lead to a stable protein conformation in the binding site. In these cases, it is important to model the missing part of the structure by computational methods. Furthermore, the stability (or instability) of conformations may be investigated in more detail by molecular dynamics simulations (see section 3.3).

### 3.2 MOLECULAR DOCKING

#### 3.2.1 Basics

The aim of molecular docking is to computationally estimate the ligand binding to a target protein. The first automated docking algorithm was introduced in 1982, which was based only on steric fitting of a ligand to the binding site, excluding all binding energetics (Kuntz et al. 1982). Today, the docking algorithms also take into account the binding energetics; in short, a small molecule ligand is fitted into a binding pocket of a protein, and the interactions between the ligand and the protein are evaluated (Fig. 5). The binding process of a ligand to a target is related to the Gibb’s energy (Eq. (1)) (Klebe 2013b), where $\Delta H$ is the change of enthalpy, $T$ is temperature and $\Delta S$ is the change of entropy.

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

The ligand–protein interactions, such as hydrogen bonding, ionic, hydrophobic, cation–π and π–π-interactions, contribute to the enthalpy of the binding energy. Based on these interactions observed in the docking pose, the docking software/algorithm produces a docking score which tries to estimate the Gibb’s energy of binding. As these interactions are the foundation of the docking, obviously virtual screening is highly dependable of the accuracy and the quantity of the data of the target system (Klebe 2006). For the entropy ($S$), certain scoring functions try to include some specific terms in order to estimate it; however, it will be mainly neglected and the docking scores are generally based on the estimation of the enthalpic term.
Figure 5 The docking procedure of fitting a small-molecule (dark-grey) into the binding pocket of a protein (light-grey) and evaluation of the interactions (dashed-lines) between the ligand and the protein. A hypothetical representation of a bad (a) and a good (b) scoring compound. With a bad-scoring compound (a), the fit is not perfect and there are less attractive interactions as compared to the good-scoring compound. An example of a docking pose of S32826 (teal; ball & sticks) docked to Autotaxin (PDB ID: 3WAX) (c). The hydrogen bonds between the ligand and the receptor are depicted as yellow dashed-lines. The binding site contains also two Zinc ions (blue, spheres), which form ionic interactions with the ligand. Glide XP-docking score for this complex is -14.788.

In the docking, the atoms and their interactions need to be modeled in a computationally efficient way such as by molecular-mechanics, with the potential energy of molecules being calculated by the force-field. In this classical mechanics approach, the parameters of the force field define potential energy function for intra- and intermolecular forces for the individual atoms, which are described as spheres in all-atom force fields. In the force field, the potential energy function is described by multiple parameters (Eq. (2)), including bonded interactions (Fig 6a–c)
and non-bonded interactions (Fig 6d) between atoms. The bonded interactions consist of bond lengths (stretching), bond angles (bending), torsions (dihedrals), out-of-plane (improper) interactions, and non-bonded interactions: van der Waals (12–6 Lennard-Jones potential) and electrostatic interactions (Coulomb). In the bonded interactions, the bond lengths ($l$), angles ($a$), dihedral angles ($\theta$) with their reference values correspond to the energy. In the distance ($r$) dependent interactions of the individual atom pairs, the depth of the energy well ($\epsilon_{ij}$), partial charges of the atoms ($q_i$ and $q_j$), permittivity of free space ($\varepsilon_0$), dielectric constant ($\varepsilon_r$) are needed to describe these terms. As noted, the force field is truly an oversimplification of the chemistry of the atoms and molecules. For instance, electrons are completely neglected. However, not only is the accuracy of a force field often sufficient to reveal various molecular phenomena, but also its simplicity enhances the speed of the computational calculations tremendously.

$$V = \sum_{i} \frac{k_{l,i}}{2} (l_i - l_{0,i})^2 + \sum_{i} \frac{k_{a,i}}{2} (a_i - a_{0,i})^2 + \sum_{i} \{ \frac{M}{k} \sum_{k} [1 + \cos(n_{ik} \theta_{ik} - \theta_{0,ik})] \}$$

$$+ \sum_{i,j} \varepsilon_{ij} \left[ \left( \frac{r_{0,ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{r_{0,ij}}{r_{ij}} \right)^{6} \right] + \sum_{i,j} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}}$$

Figure 6. Bonded interactions consist of bond length, $l$ (a); bond angle, $a$ (b); and dihedral angle, $\theta$ (c). The non-bonded interactions are described between individual atom pairs, $i$ and $j$ (d).

There are several widely used force-fields, such as Optimized Potentials for Liquid Simulations (OPLS) (Jorgensen et al. 1988, Jorgensen et al. 1996, Kaminski et al. 2001, Robertson et al. 2015); Assisted Model Building with Energy Refinement (AMBER) (Cornell et al. 1995, Hornak et al. 2006, Li et al. 2010 Lindorff-Larsen et al. 2010, Cerutti et al. 2013, Maier et al. 2015); Chemistry at Harvard Macromolecular Mechanics (CHARMM) (MacKerell et al. 1998, Best et al. 2012); and GROMOS (Schuler et al. 2001, Oostenbrink et al. 2004). The recently released version of OPLS, OPLS3 (Harder et al. 2016), has greatly enhanced the force-field’s accuracy compared to the older versions (OPLS_2005 and OPLS2.1). The functional form of the OPLS is given in Eq. (3). In particular, the torsional parameters were greatly enhanced in the update, utilizing extensive quantum chemical data. The main advantage of the force field approach is the wide coverage of
the medicinal chemistry space. In fact, the force-field’s accuracy seems to be sufficient for conducting meaningful free energy perturbation calculations, where the force-field accuracy has been one of the limiting steps of the methodology (Abel et al. 2017).

\[ E = \sum_{i<j} \left[ \frac{q_i q_j e^2}{r_{ij}} + 4 \varepsilon_{ij} \left( \frac{\sigma_{ij}^N}{r_{ij}} \right)^{12} - \frac{\sigma_{ij}^N}{r_{ij}} \right] f_{ij} + \sum_{\text{bonds}} K_r (r - r_{eq})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2 \\
+ \sum_{\text{dihedrals}} \left[ \frac{V_1}{2} (1 + \cos \varphi) + \frac{V_2}{2} (1 - \cos 2 \varphi) + \frac{V_3}{2} (1 + \cos 3 \varphi) \\
+ \frac{V_4}{2} (1 - \cos 4 \varphi) \right] \]

The molecular docking can be applied for two different purposes, in hit identification and in lead optimization. In the hit identification, a virtual library of candidate molecules is screened against the selected target. Accordingly, this process is called virtual screening. Currently, a virtual screening of a library containing several million molecules is feasible, and it can be conducted in a few weeks’ time, if reasonable computational capacity is available. In the same way as in the lead optimization, derivatives of the lead molecule are docked and evaluated. These results are utilized to guide the chemical synthesis, i.e., which molecules need to be synthesized and tested in the appropriate assay.

The ligand molecules need to be prepared before the docking (Sastry et al. 2013). The 3D-geometries of the ligands should be defined, and this is conducted by the used force-field. Moreover, the tautomers and the ionization states need to be defined. For instance, there is a crucial difference if a hydroxyl group attached to an aromatic ring exists in the –OH or in the –O– form. Obviously, the interactions and the strength of these interactions that can be formed with these tautomers are dramatically different.

Similarly, in addition to ligands, the target protein needs to be prepared. For instance, if the structural data of the protein has been obtained by X-ray crystallography, it is not ready for docking as such. One needs to add the missing hydrogens and side-chains, remove the atomic clashes, define ionization states and tautomers and conduct an energy minimization for the structure. Moreover, if the structure is missing structural data within the binding site, this also must be corrected (see section 1.1.5). Obviously, as the docking is heavily dependent on the structural data and the protein structure is usually kept rigid during the docking process, the structure preparation is not a trivial process and it exerts a major impact on the docking results (Sastry et al. 2013). Moreover, a MD simulation can be utilized to create an ensemble of protein conformations, which may aid the docking procedure and, most importantly, help to discover novel ligands towards the target (Offutt et al. 2016).

In addition, the specific ligand docking site needs to be defined prior to the docking process since without this definition, the computational effort to identify all putative interactions with a single small-molecule and the protein would be huge. In the site definition, the information of an available bound ligand or substrate can be utilized, for instance, if one would like to have the natural ligand competing with a drug, either acting as an inhibitor or an activator. However, if the ligand is absent, there are tools to identify and evaluate putative binding sites for small-molecules in the protein. The SiteMap from Schrödinger is one such tool (Halgren 2007, Halgren 2009). The SiteMap scores the putative binding sites, based on the size, functionality and solvent exposure. The tool can also be used to estimate the druggability of a target.

The docking procedure consists of pose generation for a ligand, ranking of the generated poses and finally, ranking of the docked ligands. In other words, there are two main pillars in the docking, the sampling and the scoring. First, the ligand poses are generated by the docking algorithm. This is the so-called sampling phase, which is conducted by a search algorithm. As the search for every possibly pose within the binding site for a ligand is not computationally reasonable, there are various strategies that can be applied to search for the poses for a ligand e.g.
systematic search, genetic algorithm and Monte Carlo simulation. Subsequent to pose generation, as numerous poses are generated for each ligand, the poses are ranked by a scoring function, and the best pose(s) is included in the docking output results. Similarly to the situation with search algorithms, there are various scoring functions; the three main classes of scoring function methods are a force field scoring function, an empirical scoring function and a knowledge-based scoring function (Kitchen et al. 2004). Finally, all of the docked ligands are ranked according to their docking score in the final results. In addition, one may be able to improve the ranking in some specific cases by utilizing MD simulations to rerank the docking poses (Lauro et al. 2014).

As an example of the pose generation, how the pose generation is conducted in the Glide docking algorithm is briefly described here (Friesner et al. 2004, Glide 2017). First, the initial filters test the spatial fit of the ligand to the defined active site, and furthermore, examine the complementarity of the ligand-receptor interaction using a grid-based method patterned after the empirical ChemScore function (Eldridge et al. 1997). Next, the poses that pass this initial screening state, will enter into the energy minimization stage. The poses are minimized with the OPLS force-field, and the energy-minimized poses are scored by the GlideScore. Furthermore, the poses are ranked with an Emodel score, which combines the GlideScore, the nonbonded interaction energy and the energy of the ligand conformation, and the user will receive the predefined number of the best ranking poses as the output. In other words, the ligand poses pass through a series of hierarchical filters that evaluate that ligand’s interaction with the receptor.

Currently, there are several docking algorithms and software available; such as AutoDock by the Scripps Research Institute (Morris et al. 2009), FlexX from BioSolveIT (Rarey et al. 1996) Glide (Grid-based Ligand Docking with Energetics) from Schrödinger (Friesner et al. 2004, Halgren et al. 2004, Friesner et al. 2006) and Panther (Niinivehmas et al. 2015). The most widely used docking software systems are also constantly updated. In most of the docking algorithms, the ligand is treated as being flexible while the protein structure is kept rigid. In addition, algorithms that take into account the protein flexibility to some extent exist, and allow some degree of flexibility in the protein; for example, Induced Fit Docking from Schrödinger (Sherman et al. 2004, Farid et al. 2006, Sherman et al. 2006) and PLANTS (Korb et al. 2006, Korb et al. 2007). However, the drawback of these flexible-protein methods is that they are computationally expensive, and therefore limited to only selected ligands (up to hundreds to thousands). Also, the ranking of poses may be encumbered by the scoring of different conformations. This is dependent on how well the scoring function able to score different conformations in parallel.

3.2.2 Analysis

Validation of docking is usually conducted with known active molecules. This is conducted prior to the screening or docking. In validation, a set of known active molecules (if available) should produce good poses with a high-rank in the docking score with a decoy-set of inactive molecules. For instance, a decoy set of molecules can be generated based on the existing set of ligands (Mysinger et al. 2012) This can be analyzed by the enrichment factor, which is the true binder enrichment among the top-ranking poses with the decoy set of inactive molecules (Huang et al. 2006). The enrichment factor is, however, highly-dependent on the number of active molecules; therefore, it tends to prevent a valid comparison with different datasets. There are alternative methodologies to analyze the enrichment, i.e. the receiver operating characteristics (ROC) curve (Triballeau et al. 2005) or the more advanced Boltzmann-enhanced discrimination of receiver operating characteristics (BEDROC) (Truchon and Bayly 2007). These ROC methods result in a plot where an area under the curve (AUC) value reveals the extent of the enrichment. The AUC value of 1 represents perfect enrichment whereas a value of 0.5 represents a random value. However, both of these two methods require that there is data of known active molecules in addition to their specific binding site, and can be considered rather as “hindsight”-methods. Therefore, these methods are virtually useless if one does not have information of the known binders and their specific binding site. This can be the case, especially, when novel targets are identified and validated, and there exists no data about the binding site(s) and/or ligand(s).
Although the docking score produced by the attempts of the used docking software/algorithmt to estimate the binding energy, it is still a somewhat arbitrary value. Moreover, some scoringfunction usability is dependent on the target; in other words, one scoring function can performbetter than some other with a specific target, whereas the situation could be reversed for someother target (Warren et al. 2006). Therefore, also the visual analysis of the docking poses is crucialwhen evaluating the poses. The docking score can be, however, used as a good general guidancein the selection process. In other words, the million-molecule input can be filtered according tothe docking score, i.e. it will reduce the scrutiny down to the best scoring hundred(s) of molecules.These best molecules can be visually evaluated, and discarded or selected for further analysis orcan be tested in a suitable in vitro assay. Naturally, in the selection process, other chemicalproperties of the candidate molecules can be also used as selection criteria. For instance, themetabolic susceptibility of the potential candidate molecule may be considered at this juncture.

In fact, for further validation of a docking pose, a MD simulation can be conducted (Cavalli etal. 2004, Alonso et al. 2006, Colizzi et al. 2010, Sabbadin et al. 2014, Bartuzi et al. 2017). Duringthesimulation, if the pose appears stable, it is an indication of the validity of the pose. In addition,asteered-MD method which evaluates the ligand dissociation and thereby the docking validitywas recently introduced (Ruiz-Carmona et al. 2017). However, these methods are, obviously,unsuitable for handling thousand(s) of molecules.

Obviously, the best feedback for the docking is obtained, when the selected molecules aretested in a relevant assay. In the assay, the true hits are identified, and the feedback, dependingon the obtained hit-rate, is self-explanatory. The hit-rate (%), which is the true hits (active in theassay) divided by the total number of tested molecules, depends greatly on the target. Therefore,it cannot be utilized to evaluate the effectiveness of the docking. For instance, in the search forPTP-1B inhibitors, a hit-rate of 34.8% (IC\textsubscript{50} < 100 µM) for the tested 365 molecules was obtained(Doman et al. 2002), whereas in our Autotaxin VS campaign, two of the tested 26 compoundsshowed inhibition of ATX (tested in 10 µM), i.e. the hit-rate was 7.7% (Pantsar et al. 2017).Moreover, based on the feedback from the assays, one is able utilize this information to furtherenhance the docking.

### 3.2.3 Limitations

It has to be appreciated that the docking is only a snapshot of the binding event. Therefore, asstated above, the quality of the structural data cannot be emphasized enough. With a poor-qualityprotein structure, one is not able to achieve and should not expect to obtain useful docking results.Here, how the structure was obtained, plays a crucial role (see section 3.1). In addition, thecharacteristics of the binding site are crucial for the docking. If the binding site is highly-dynamic,itis extremely difficult to achieve good docking results. As the protein conformation may changeDRAMATICALLY after the binding of the compound, the results obtained with rigid docking aretotally arbitrary. Basically, the protein conformation and the absence of the temporal-aspect arethe key-issues with molecular docking. Furthermore, even though the issues with the uncertaintyof the data would not exist, the ability of the scoring functions to estimate the binding affinity isextremely difficult. This is due the fact that the docking is based on the force-field, which causeapproximation error to the process. By applying quantum mechanics this error could be reduced;however, this is usually computationally too expensive. Finally, the entropy (S), which is animportant factor in the binding process (Eq. (1)), is neglected with docking approach, and thereis currently no computational technique in docking available to estimate it extensively (Procacci2016).
3.3 MOLECULAR DYNAMICS

3.3.1 Basics
The molecular dynamics (MD) simulations reveal the spatiotemporal behavior of the studied system at the atomic scale. In other words, a molecular dynamics simulation generates a trajectory where the individual atom movement is observed as a function of time. Not only can MD simulations provide information of the studied biomolecular system (Dror et al. 2012), but also their valuable role in drug discovery has emerged lately (De Vivo et al. 2016). There are multiple applications available, these include ligand binding kinetic analysis, the combination with molecular docking methods (see section 3.2) and the solvent analysis (see section 3.4).

The classical all-atom molecular dynamics are based on two main pillars. First, the movement of the atoms are governed by the laws classical mechanics i.e. the Newton's second law. Second, the atoms in the system are described by a force field (see section 3.2.1). In the Newton's second law (Eq. (4)), the net force \( f_i \) for an atom \( i \) at time \( t \) is equal with its mass \( m_i \) and its acceleration \( a_i \). This corresponds to the empirical potential energy function \( V(x) \), which is described by the force field.

\[
f_i(t) = m_i a_i(t) = -\frac{\partial V(x(t))}{\partial x_i(t)}
\]

Currently, there exist various molecular dynamics simulation software packages with the most widely used being GROningen Machine for Chemical Simulations (GROMACS) (Berendsen et al. 1995, Abraham et al. 2015); NAMD (Phillips et al. 2005), which can be applied with GUI via VMD (Humphrey et al. 1996); Amber (Pearlman et al. 1995, Case et al. 2005, Case et al. 2017), Chemistry at HARvard Macromolecular Mechanics (CHARMM) (Brooks et al. 1983, Brooks et al. 2009) and Desmond, which can be combined with Schrödinger’s Maestro GUI (Bowers et al. 2006, Guo et al. 2010, Shivakumar et al. 2010, Desmond Molecular Dynamics System 2017).

The limitations of classical MD simulations consist of the pre-defined parameters, which are basically related to the force-field (De Vivo et al. 2016). First, the tautomers and ionization states are defined prior to the simulation, and these do not change during the simulation. Furthermore, the force-field defined fixed charges with the atoms are an issue, as naturally there exists polarizability in specific environments for the charges e.g. in the protein–ligand binding events. However, polarizable force fields have been developed (Baker et al. 2015). Unfortunately, these force fields come at an extensive computational cost, which limits their current usability. Most importantly, the existing bonds between the atoms cannot break and new bonds cannot form during a classical MD simulation. This can be partially addressed with a hybrid quantum mechanics molecular mechanics (QM/MM) simulation, but in reality, this is applicable only to defined site(s), such as a catalytic site (Warshel & Levitt 1976, Senn & Thiel 2009, Mones et al. 2015, Torras 2015). Moreover, the simulations are not currently capable of handling the natural composition of a biological systems, because of the numerous different molecules present in natural systems (Dror et al. 2012)

3.3.2 Timescale and protein movement
The timescale is one important aspect which needs to be taken into account when studying protein dynamics with MD simulations. Different types of protein movement occur at different timescales (Fig. 7) (Henzler-Wildman and Kern 2007). Side-chain rotamers are observed in ps to μs timescales, loop motions in ns to μs timescales, where larger domain motions appear only in the μs to ms timescale. Therefore, in practice, this means that one needs to have a sufficient amount of simulation data if one wishes to observe or assess the motions of interest. For example, if one wishes to observe larger domain motions of the protein, a simulation consisting of hundreds of nanoseconds will be totally insufficient.
To assess the occurrences of these events reliably, not only does one need to have a long enough simulation, but also replicates of the simulation. These replicate simulations, with slightly different starting configurations, reduce the chance to overanalyze a random observation from a simulation. Moreover, a crucial issue in the simulations is the sampling, i.e. does a simulated system visit all important conformational states during the simulation. Usually there is enough sampling for the faster local motions. However, to obtain adequate sampling especially for the systems with slow relaxations may be problematic (Romo and Grossfeld 2014). The reliability of a simulation (i.e. what is observed in a simulation) is also affected by the initial system setup (e.g. protein conformation) and the force-field accuracy. To address the sampling issues, different enhanced sampling methodologies have been developed (Harpole and Delemotte 2018). For instance, replica-exchange (Sugita and Okamoto 1999) and simulated tempering (Pan et al. 2016).

![Figure 7. Timescale of dynamic processes in protein. The local flexibility occurs in the fs to ns timescale, whereas collective motions occur in the µs to ms timescale. MD simulations cover the fs to ms timescale (Adapted from Henzler-Wildman and Kern 2007).](image)

In 2007, the longest conducted MD simulation with a protein structure was only 2 µs (Ensign et al. 2007). Therefore, a mere ten years ago, the information obtained from MD simulations was extremely limited. However, substantial progress has been made in the field, and finally, a simulation of over 1 ms was introduced in 2009 (Shaw et al. 2009, Shaw et al. 2010). In comparison, our conducted MD simulations of K-Ras had an MD simulation time of 170 µs (Pantsar et al. 2018). On the whole, a sufficient simulation time to make it possible to study also larger domain motions in proteins is accessible nowadays with the current MD simulations.

### 3.3.3 Analysis

The data generated by an individual simulation is enormous. For example, a simulation of a small globular protein of 167 residues in a water box (radius 13 Å from the protein), comprises approximately 33 000 individual atoms (the protein, ligand, solvent and ions). If the system is simulated for 2 µs time, and the atomic coordinates are saved at 2 ps intervals, the generated trajectory would have 1 000 000 frames. This results in a total of 33 000 000 000 atomic coordinates (number of atoms multiplied by frames). Naturally, one may not be interested in the locations of the atoms at every second picosecond; therefore, the coordinate saving interval can be reduced. Nevertheless, in most cases, the system needs to be replicated to minimize random bias, and most likely a longer simulation time is needed, resulting in a tsunami of data.
The root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) are good first-line analysis tools for analyzing a molecular dynamics simulation. In both approaches, a selected structural element of the system, usually the protein or the small-molecule ligand, is compared to a reference structure; for instance, the input crystal structure used as the starting conformation of the simulation. In RMSD, the deviation of the selected atomic coordinates (e.g. backbone) from the reference structure is monitored as a function of the simulation time. In contrast, the RMSF provides information about the overall fluctuation of individual residues or atoms in the whole simulation, or within in a specifically selected timeframe. Moreover, these tools reveal more about the validity of the simulation itself, and they do not provide true insights into the simulation data. In addition, assessment of distances and monitoring hydrogen bonds or other interactions are easy and straightforward ways to analyze a system. It is straightforward to measure these parameters between selected individual atoms, or whole residues; especially, if some specific interaction(s) of interest is known.

If one wishes to acquire more insight into the dynamics, one needs to apply mathematical methods to filter the vast MD simulation data. One method to filter noise from the data is the principal component analysis (PCA), also called covariance analysis or essential dynamics (Amadei et al. 1993). This is a multivariate statistical technique, which identifies the largest global correlated motions of the protein, thereby reducing the dimensions of the simulation data. First, rotation and translation of the protein are removed by fitting with a reference structure. Next for the dimension reduction, a linear transformation is conducted, where a covariance matrix is applied for the simulation data (David & Jacobs 2014). The covariance matrix is usually constructed based on a coarse-grained description of the protein; $\text{Ca}$ atom coordinates or $\text{Ca}$ torsion angles are commonly used for the matrix generation, resulting in a symmetric $3N \times 3N$ matrix. Eigenvalue decomposition of the covariance matrix results in the production of a complete set of orthogonal collective modes, the eigenvectors. The variance of these eigenvectors, called eigenvalues, characterizes the share of the corresponding motion. The principal component is the resulting projection of the simulation data onto an eigenvector.

The time-lagged independent component analysis (TICA) is similar linear transformation dimension reduction method as PCA, but it also identifies the slow order parameters of the simulation (Peréz-Hernández et al. 2013, Schwantes & Pande 2013). In order to achieve this, TICA maximizes the autocorrelation of the transformed coordinates in the linear transformation, and thereby it finds a maximally slow subspace. In TICA, one can use various parameters for input (coordinates, distances, torsions, of selected atoms), but also a lag time needs to be defined. In fact, TICA is an appropriate method to prepare data for Markov model construction (Peréz-Hernández et al. 2013). TICA is an ideal approximation to the Markov operator’s eigenvalues and eigenfunction, and thereby it can be utilized with Markov state model (MSM) generation, which is a relatively new MD analysis method.

In fact, in the near future, a MSM could be one of the most important tools to analyze molecular dynamics simulations. Accordingly, in recent years, the MSMs have been utilized in a growing number of MD analyses, including protein conformational dynamics (Da et al. 2016, Wieczorek et al. 2016, Gao & Zhao 2017), ligand binding kinetics (Plattner & Noé 2015, Casasnovas et al. 2017) and protein–protein association kinetics (Plattner et al. 2017).
MSM is a kinetic network model, which identifies metastable states of the system and estimates probabilistic transitions between these states (Fig. 8). Currently, there are freely available open source software packages for MSM creation, which make model creation relatively straightforward. The two most widely used software systems are PyEMMA (http://emma-project.org/latest/) and MSMbuilder (http://msmbuilder.org/3.7.0).

Figure 8. A hypothetical representation of a Markov State Model’s metastable states and the state transitions. The size of an oval is related with the probability of a state existence at time $p_{i0}$. The thickness of an arrow represents the state transition probability.

In a MSM generation, simulation data from a few long simulations or multiple short simulations is needed. First, a dimension reduction is required for the simulation data, which is usually conducted with TICA. In addition, the lag time for the model should be defined. Next, the structures in MD trajectories are assigned into microstates. These microstates are further coarse-grained into macrostates; for example, with Perron-cluster cluster analysis (Roeblitz & Weber 2013). These macrostates represent the metastable states, and the corresponding conformations can be generated for each state. The output from MSM is the metastable states and the state probabilities. Moreover, the state transition timescales can be estimated.

As MSM is a model, a validation of the generated MSM is always needed. First, the implied timescales need to be constant at the lag time used. This indicates the markovianity of the generated model. Furthermore, as the implied timescales are not alone a sufficient indication of the model validity, a Chapman-Kolmogorov test needs also be conducted. This test is a good indication about how well the predicted model is able to follow the observed trajectory. One limitation of the MSM is that a sufficiently long simulation time is needed in order to generate a valid model. In most of the published MSM studies, the overall simulation time of an individual system has been in the range of tens to hundreds of microseconds.

To summarize, a series of short simulations (simulation timescale hundreds of nanoseconds) may provide useful insights, not only for drug binding process, but also about plausible effects of a small-molecule on the drug target. For instance, with a ligand, some regions of the protein may start to display specific dynamics compared to the reference dynamics. Moreover, if one metastable state can be identified to be related to a protein–protein interaction (binding of another protein), then the shift in dynamics away from this metastable state may lead to decreased binding with this binding partner.
3.4 HYDRATION SITE ANALYSIS

The water molecules within the binding pocket play a significant role in small-molecule binding to a protein. Usually, the binding site is occupied with water molecules and the displacement of these water molecules by the small-molecule ligand is the principal source of the binding energy (Abel et al. 2008). These water molecules within the binding site are energetically either favorable or unfavorable compared to the bulk water. Most often the water molecules lose their freedom to some degree as they are more or less bound to the binding site, and they are surrounded by protein residues and not the bulk solvent. Therefore, there can exist an entropic penalty for a particular water molecule binding within the site. In addition, there can exist an enthalpic penalty, if the water molecule is not able to form hydrogen bonds in the site; for example, in the proximity of lipophilic side-chains. If the total energy of the water molecule is positive (unfavorable), and this kind of water molecule is displaced from the site, there will be an energetic gain and this promotes the small-molecule binding to the site. On the other hand, if an energetically negative water molecule is displaced from the site, it will lead to a penalty in the binding, and a decrease in the binding affinity. Moreover, if the ligand binding will trap water molecule(s) and transform the binding energy of these waters so that it becomes unfavorable, it will also decrease the binding affinity of the ligand.

There exist software that can evaluate the hydration sites on the binding cavity in silico (Bodnarchuk 2016), and thereby aid in the small-molecule design process e.g. 3D-RISM in Molecular Operating Environment (MOE) (Kovalenko et al. 1999, Luchko et al. 2010), and the WaterMap from Schrödinger (Young et al. 2007, Abel et al. 2008, WaterMap 2017). The WaterMap software conducts a short molecular dynamics simulation, and based on that information, it estimates the energies of the hydration sites (Fig. 9). This estimation can be conducted with or without the ligand, as naturally, the ligand can exert a dramatic effect on the hydration network in the site. These hydration site energies can be thereafter utilized in the ligand design. In other words, it reveals which hydration sites should be occupied with the ligand and which sites should be left intact, correspondingly, in order to gain or not to lose binding affinity. However, this short simulation is highly-dependent on the conformation of the protein. In that respect, if the used protein conformation is arbitrary for the studied ligand–protein complex, the results of the WaterMap analysis will be invalid.
3.5 PROTEIN–PROTEIN INTERACTION PREDICTION IN SILICO

Currently, there are several protein–protein docking software systems; however, mostly only the binding prediction of targets with large interfaces show reasonable performance, whereas with others, poor performance is likely (Gromiha et al. 2016, Lensink et al. 2016). Therefore, we utilized the GRID method to predict phosphorylated Myc’s phosphate binding site to AurkA (Goodford 1985, Dauch et al. 2016). The GRID method determines energetically favorable binding sites on a macromolecule. In the method, an interaction of a probe group is computed at sample positions throughout and around the macromolecule, producing an array of energy values. The probes include water, the methyl group, amine nitrogen, carboxyl oxygen, and hydroxyl. The resulting contour surfaces can be visualized based on their energy values.

3.6 METHODS TO ANALYZE PROTEIN–PROTEIN INTERACTIONS IN VITRO

The ability to block a protein–protein interaction of the drug target and a specific interaction partner, can be a determining aspect of a small-molecule’s pharmacodynamic efficacy. Therefore, it is crucial to identify if a developed small-molecule, in addition to its ability to bind to the drug target, is able to prevent the crucial protein–protein interaction. There exist a plethora of biochemical and biophysical methods to analyze protein–protein interactions. In this section, only the methods with the highest relevance to the conducted studies will be briefly described. Moreover, the methods to obtain the structural data (described previously; see section 3.1) can be utilized in specific cases to study the protein–protein interactions. Finally, a combination of different methods may be useful in selected cases.
3.6.1 Isothermal titration calorimetry
In isothermal titration calorimetry (ITC), a solution of a protein or a small molecule is titrated in a solution with the target protein (Klebe 2013a). The binding event may be an endothermic or an exothermic process. Accordingly, it generates a heat-signal that can be measured. The free-energy of binding (ΔG) is obtained from this measurement, and in addition, individual proportions of the enthalpy (H) and entropy (S) can be determined. Subsequently, the affinity can be derived from the results. Furthermore, ITC provides information about the stoichiometry of the binding.

The method is widely utilizable with various protein–protein interactions. Moreover, as it is a label-free method, it reduces the risk for artificial results. There are numerous publications available where ITC has been utilized e.g. protein-DNA interaction (Tzeng & Kalodimos 2009), integrin complex formation in phospholipid bicelles (Situ et al 2014) and microtubule-associated proteins - tubulin (Tsvetkov et al. 2013). The drawback of the method is that large amounts of protein(s) are needed for the measurement and the solubility may cause issues.

3.6.2 Surface plasmon resonance
The surface plasmon resonance (SPR) method is based on a change in the light refractivity index shift upon binding (Schuck 1997, Löfås 2004). A protein (or a small-molecule) is attached to a metallic sensor, for example to a gold-plate chip. For instance, this can be achieved after the biotinylation of the protein, which is attached to streptavidin on the chip. A flow of the other studied protein (or a ligand) passes adjacent to the chip. The light excites surface plasmons in the metal, and if binding occurs, this causes mass changes on the chip, leading to alterations in the surface plasmon signal. The binding shifts the refraction of light from the plate and the binding is detected.

The main detection formats used with SPR are direct detection, sandwich detection, competitive detection or inhibition detection (Homola 2008). In direct binding, there is a single binding event of the analyte to the attached protein. In sandwich detection, there is an additional binding event after the initial binding. A large molecule with analyte conjugate is used in the competitive binding detection that competes with the analyte. The inhibition detection involves a prior incubation with the chip-bound element (without the chip). Next, in the flow phase, only the unbound fraction of target will bind to the chip-element, whereas occupied target proteins will pass by and not result in a signal.

The advantage of the SPR method is that even a small/weak binding can be detected. Moreover, because of the sensitive detection, only a low-amount of protein is needed. It is also a label-free method, except for the fact that one element need to be attached to the chip, which usually requires some modification (e.g. biotinylation). In some cases, this required modification may result in issues with the system setup; thus, making SPR unusable method with some targets. In addition, the target residence-time of a ligand can be measured with the method (Copeland 2016, Willemsen-Seegers et al. 2017).

3.6.3 Coimmunoprecipitation
Coimmunoprecipitation is a biochemical method, which identifies protein–protein interactions that exist in a cell (Adams et al. 2002, Nature Methods 2005). In the procedure, the cells are lysed and the protein of interest is immunoprecipitated, i.e. captured by a specific antibody. If the protein of interest undergoes an interaction with another protein, they will both coimmunoprecipitate together. Obviously, the conditions in the procedure need to be nondenaturating. Next, the coimmunoprecipitated proteins are detected by autoradiography or western blotting. The method can be used to identify if a specific protein–protein interaction exists in a particular setting, or to identify the existing interactions with a studied protein. One crucial and the limiting aspect of the method is that one needs to have a suitable antibody for the target protein. The antibody needs to be selective and to exhibit high-affinity against the target.
3.6.4 Kinase activity assays

In the kinase activity assays, the goal is to monitor the inhibition of the kinase activity of the target protein kinase (Dar & Shokat 2011). However, as phosphorylation is a crucial step in this protein–protein interaction signaling transduction, with a kinase activity assay one is able to estimate indirectly the end-point of this type of protein–protein interaction. Therefore, this section is included here, among the methods of analyzing protein–protein interactions. It is noteworthy that some of the methods described in this section can also be utilized as a direct protein–protein interaction measurement; e.g. Fluorescence polarization (FP), enzyme-linked-immunoabsorbent assays (ELISA), Fluorescence resonance energy transfer (FRET) and Luminescent oxygen channeling (Arkin et al. 2012). This section tries to highlight how the current kinase activity assays reflect the in vivo situation, where there exists a protein–protein interaction (or multiple interactions) prior to the target protein phosphorylation.

There are a plethora of protein kinase inhibition assays in the literature (Glickman 2012). These include direct methods, which measure the target substrate phosphorylation i.e the reaction product, and indirect methods that do not detect directly the phosphorylated product, but estimate the inhibition in some indirect manner.

FP(anisotropy) based methods measure the change in the anisotropy caused by the change of the Brownian motion of the target fluorophore (Checovich et al. 1995). The relatively small fluorescently-labeled substrate peptide binds to an antibody after its phosphorylation, i.e. the antibody needs to be specific for the phosphorylated substrate (Parker et al. 2000). Moreover, the methodology can be used without the antibody, by utilizing an immobilized metal ion which binds to the phosphorylated substrate (Turek-Etienne 2003). The downside of the anisotropy methods is that the fluorescence assays are susceptible to assay interference (intrinsic fluorescence of the compounds).

In a Fluorescence resonance energy transfer (FRET) assay, the energy transfer from an excited donor fluorophore to an acceptor fluorophore is detected (Rodems 2002). The target peptide is labeled e.g. with fluorescein (donor) and coumarin (acceptor). Moreover, a protease which is able to degrade the target peptide is needed for the assay. The phosphorylation event to the target peptide, in the vicinity of the cleavage site, reduces the ability of a protease to cleave the target peptide, thereby affecting the detected signal. The method is susceptible to assay interference from protease inhibitors. Moreover, since it is a non-direct method, it is also able to identify allosteric inhibitors. However, one can not distinguish the inhibition type based on the assay.

Homogenous time resolved fluorescence (HTFR) (also known as Time Resolved Förster Resonance Energy Transfer (TR-FRET)), detects the energy transfer between Europium cryptate or Terbium cryptate and an acceptor fluorophore, such as allophycocyanin (Degorce et al. 2009). An antibody, which identifies the phosphorylated substrate, is labeled with the donor cryptate. Moreover, the substrate is also tagged with an element that binds the acceptor fluorophore; for instance, the recognition element can be streptavidin–biotin or an antibody. The method can also be used to analyze kinase autophosphorylation (Moshinsky et al. 2003).

In the Scintillation proximity assay (SPA), the phosphorylation of the substrate is detected by the β-emission of the radioactive 33P isotope. When the the scintillant bead is bound to the phosphorylated (33P) substrate, the energy transfer is conducted to the bead, causing it to emit light (Cook 1996). As this β-emission travels only a limited distance in a solution, the unbound 33P (e.g. ATP) does not result in signal. Obviously, the substrate needs to be able to bind to the beads; for instance, the kinase substrate can be biotinylated, which binds to the streptavidin labeled beads (Park et al. 1999). The method is relatively free from assay interference (Glickman et al. 2008). However, the need for a radioactive material is a disadvantage of this method due to the cost and handling issues.

In the luminescent oxygen channeling (Ullman et al. 1994), a laser is used to excite the donor bead, producing singlet oxygen. This singlet oxygen is able to diffuse 200 nm in solution. Therefore, if the acceptor bead lies in that range, i.e. in the close proximity of the donor, it results in the production of light. The antibody, which identifies the phosphorylated product, is utilized...
to bring the donor and acceptor beads into close proximity. In other words, only the phosphorylated product will result in the detected signal. The method is suitable for high-throughput screening (Von Leoprechting et al. 2004).

Furthermore, an immunoabsorbent assay can be utilized to assess kinase inhibition. Usually, these are enzyme-linked-immunoabsorbent assays (ELISA) (Zhang et al. 2007a). In an ELISA assay, a specific antibody that is able to distinguish phosphorylated substrate is added, and either this or a secondary antibody is linked to an enzyme, which produces the signal after addition of its substrate.

Luciferase-based ATP detection assay measures ATP consumption (Koresawa & Okabe 2004). In the assay, the quantity of ATP (remaining) is measured with luciferase enzyme, where the enzyme needs ATP for the oxidation reaction of luciferin (in the reaction ATP is transformed to AMP), which simultaneously produces luminescence as a byproduct. It is a fast, high-throughput method. The method is susceptible to interference with the compounds that cause luciferase inhibition.

DiscoverX has developed competition based kinase assays, for example, the Enzyme Fragment complementation which utilizes the β-galactosidase enzyme, with the competitor ligand being linked to a fragment of the enzyme. (Vainshtein et al. 2002). The unbound competitor ligand binds to the inactive incomplete β-galactosidase, resulting in an activated enzyme. The signal obtained from this enzyme reaction in the competition assay is directly proportional to the ligand concentration. Subsequently, a more modern competition assay without the need of an enzyme has been developed (Fabian et al. 2005). The bait-ligand is tagged with biotin and, furthermore, bound to a solid surface. Next the kinase and the unlabeled native test inhibitor is placed into the competition with the bait-ligand. The kinase protein is tagged with T7 phage particle, thereby the amount of kinase in the solid surface, which is bound to the bait-ligand, is quantified with phage plaque assays or quantitative PCR (with phage DNA template). From the extent of the competition, it is possible to calculate the affinity of the tested ligand and moreover, the kinase inhibition. The method is also able to detect ATP-adjacent allosteric inhibitors, which alter the conformation of the ATP-binding site. However, it is not likely that allosteric inhibitors which do not affect to ATP binding will be detected. Large selectivity screens have been conducted with this method (Davis et al. 2011).

The mobility-shift kinase activity assays are based on the detection of the phosphorylated substrate. As the target substrate peptide is phosphorylated, the net-charge of the peptide is changed (the phosphate’s negative charge) and any change in the substrate’s net-charge will result in a mobility shift. The mobility shift assays can be utilized for instance, in kinase activity profiling from single cell lysates (Cheow et al. 2014), as well as for kinase selectivity profiling (Drueckes 2016).

Moreover, the binding affinity of a compound to the kinase itself, for example conducted with a thermal shift assay (Lo et al. 2004), seems to show a good correlation with the kinase inhibition (Rudolf et al. 2014). This is relevant when the affinity assay is conducted with full-length kinases and not just with the catalytical kinase domain.

As can be seen from above, it is important to understand what type of an assay is being utilized and what type of kinase inhibitors it may be able to identify. For instance, the competitive assays are not able to detect allosteric inhibitors. It is important to note that the inhibition of kinase activity is not always directly linked to the inhibitory properties for the signaling activity (Dar & Shokat 2011), a fact that will be emphasized in the Aurora kinase A chapter.
4 Case Studies of Selected Drug Targets

In this chapter, three dissimilar drug targets Aurora kinase A (AurkA), Autotaxin (ATX) and small GTPase protein K-Ras (K-Ras), are discussed from the perspective of drug design and particularly, their utilization against the solid tumors for which there are insufficient treatment options. Finally, the drug design on these targets are elaborated from the perspective of the efficacy problem related with each target.

4.1 AURORA KINASE A

4.1.1 Aurora kinase A as drug target

Various members of the protein kinase superfamily are dysregulated in cancer (Fleuren et al. 2016). This is not surprising, as the protein kinases are the key regulators of many cellular processes and most of the cellular signaling pathways are regulated by protein phosphorylation (Endicott et al. 2012). A kinase protein, catalyzes the $\gamma$-phosphate transfer from an ATP molecule to a target protein’s tyrosine, serine or threonine residue; this regulates the target protein’s activity, usually leading to its activation or inactivation. AurkA is a serine/threonine kinase, i.e. AurkA phosphorylates serine or threonine residues of its target proteins. AurkA is a 45.8 kDa protein, comprising 403 residues. The protein kinase domain of AurkA consists of the residues 133–383. AurkA has no isomers and the encoding gene is located in chromosome 20. In the human kinome, AurkA belongs to the group of Other kinases. In the aurora kinase protein family, two other members exist; Aurora kinase B (AurkB) and Aurora kinase C (AurkC) (Fig. 10). These proteins are slightly smaller in size, and they display a high-difference with AurkA in their N-terminal part. Thus, although highly similar in the other parts, AurkB and AurkC exhibit only 51% and 49% overall identity (calculated with Clustal Omega; Goujon et al. 2010, Sievers et al. 2011) with AurkA, respectively (Li et al. 2009). Moreover, the other family members have distinct cellular functions compared to AurkA, such as chromosome condensation, chromosomal alignment and segregation (Tang et al. 2017a).
AurkA has a rather typical structure for a kinase (Fig. 11). The structure consists of an N-lobe and a C-lobe, where the ATP-binding site is located in the deep cleft between these lobes. There have been numerous, almost one hundred, crystal structures of AurkA published to date. However, the structural data of the protein is incomplete, as the structures include usually residues ~120–400; thus, lacking quite a large proportion of information of the N-terminal part of the protein. The exact structure of this N-terminal region is unknown to date; however, it is believed to participate in the regulation of AurkA’s activity by autoinhibition via an interaction with the C-lobe (Zhang et al. 2007b, Bai et al. 2014) and to the localization (Giet & Prigent 2001).
AurkA phosphorylates numerous target proteins (Table 1). These include different kinases; proteins from Ras-signalling pathway; well-known tumor suppressors, such as BRCA1 and p53; and various proteins involved in cell-division.

Not only does AurkA regulate other proteins via its kinase activity, but also AurkA itself is regulated by phosphorylation. AurkA is regulated by phosphorylation in five positions: S41, S51, T287, T288, S342. The AurkA was shown to be phosphorylated at S41 during M-phase (Daub et al. 2008), but the role of this phosphorylation is still unclear. The phosphorylation of S51 inhibits AurkA’s proteasomal degradation (Littlepage & Ruderman 2002) and, thereby, the dephosphorylation of S51 by Protein phosphatase PPA2 controls AurkA’s degradation (Horn et al. 2007). The atypical protein kinase C (aPKC) phosphorylates AurkA at T287, which augments AurkA’s binding with Targeting protein for Xklp2 (TPX2) and enhances activation of AurkA (Mori et al. 2009). The phosphorylation of T288 leads to the activation of AurkA, as T288 is located in the activation loop. This phosphorylation results in a change of the dynamics/conformation of the activation loop, resulting in the access and binding of the protein substrate. Protein kinase A (PKA) and Serine/threonine-protein kinase PAK1 (PAK1) are able to phosphorylate T288, and activate AurkA (Walter et al. 2000, Zhao et al. 2005). Moreover, AurkA can autophosphorylate T288, when bound to TPX2 (Zorba et al. 2014). In the complex with TPX2, AurkA forms a catalytically active dimer, where the autophosphorylation takes place intermolecularly. In addition, TPX2 promotes AurkA’s active conformation and stabilizes a water-mediated allosteric network, which activates AurkA (Cyphers et al. 2017). Furthermore, the binding with TPX2
shields it from deactivating dephosphorylation of T288 (Bayliss et al. 2003). The protein phosphatase 1 (PP1) can dephosphorylate T288 and thereby inhibit AurkA (Walter et al. 2000). Interestingly, AurkA itself is able to phosphorylate PP1 and inactivate it during mitosis (Katayama et al. 2001). PAK1 is also able to phosphorylate AurkA at S342, which also enhances kinase activity of AurkA, albeit to a lesser extent than the phosphorylation at T288 (Zhao et al. 2005). Overall, the highest phosphorylation status of AurkA is observed during the M phase in the cell-cycle.

Table 1 Target proteins that AurkA phosphorylates.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Full name</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARHGEF2</td>
<td>Rho guanine nucleotide exchange factor 2</td>
<td>Birkenfeld et al. 2007</td>
</tr>
<tr>
<td>BORA</td>
<td>Protein aurora borealis</td>
<td>Hutterer et al. 2006, Seki et al. 2008</td>
</tr>
<tr>
<td>BRCA1</td>
<td>Breast cancer type 1 susceptibility protein</td>
<td>Ouchi et al. 20041, Ertych et al. 2016</td>
</tr>
<tr>
<td>CPEB1</td>
<td>Cytoplasmic polyadenylation element-binding protein 1</td>
<td>Sasayama et al. 2005</td>
</tr>
<tr>
<td>CDC25B</td>
<td>M-phase inducer phosphatase 2</td>
<td>Dutertre et al. 2004</td>
</tr>
<tr>
<td>DAP-5; HURP</td>
<td>Disk large-associated protein 5</td>
<td>Yu et al. 2005</td>
</tr>
<tr>
<td>HDAC6</td>
<td>Histone deacetylase 6</td>
<td>Pugacheva et al. 2007</td>
</tr>
<tr>
<td>LATS2</td>
<td>Serine/threonine-protein kinase LATS2</td>
<td>Toji et al. 2004</td>
</tr>
<tr>
<td>NDEL1</td>
<td>Nuclear distribution protein nudE-like 1</td>
<td>Mori et al. 2007</td>
</tr>
<tr>
<td>PARD3</td>
<td>Par polarity complex protein Par3</td>
<td>Khazaei and Püschel 2009</td>
</tr>
<tr>
<td>PLD2</td>
<td>Phospholipase D2</td>
<td>Mahankali et al. 2015</td>
</tr>
<tr>
<td>PLK1</td>
<td>Polo-like kinase-1</td>
<td>Macůrek et al. 2008</td>
</tr>
<tr>
<td>RaLA</td>
<td>Ras-related protein RaL-A</td>
<td>Wu et al. 2005</td>
</tr>
<tr>
<td>RASSF1A</td>
<td>Ras association domain-containing protein 1A</td>
<td>Rong et al. 2007</td>
</tr>
<tr>
<td>TACC3</td>
<td>Transforming acidic coiled-coil-containing protein 3</td>
<td>LeRoy et al. 2007</td>
</tr>
<tr>
<td>TP53; p53</td>
<td>Cellular tumor antigen p53</td>
<td>Katayama et al. 2004</td>
</tr>
<tr>
<td>TPX2</td>
<td>Targeting protein for Xklp2</td>
<td>Kufer et al. 2002</td>
</tr>
</tbody>
</table>

† The article was subsequently withdrawn

Furthermore, AurkA displays additional interactions with proteins, for which it does not display any kinase activity. These interacting proteins include NAD-dependent protein deacetylase sirtuin-2 (SIRT2) (North and Verdin 2007), LIM domain-containing protein ajuba (AJUBA) (Hirotat al. 2003), Protein pitchfork (PIFO) (Kinzel et al. 2010), Enhancer of filamentation 1 (hEFl, also known as HEF1; gene NEDD9), RAF proto-oncogene serine/threonine-protein kinase (CRAF, also known as Raf-1) (Mielgo et al. 2011), Calmodulin (Plotnikova et al. 2012) and, also the transcription factors N-Myc (Richards et al. 2016) and Myc (Dauch et al. 2016). AurkA regulates the proteasomal degradation of these transcription factors, as it shields them from the ubiquitin ligases. Therefore, the ability to regulate the activities of other proteins by Aurka is not only limited to its kinase activity. Moreover, growth arrest and DNA damage-
inducible protein GADD45 alpha (GADD45A), which is regulated by p53 and BRCA1, are able to bind AurkA and inhibit its kinase activity (Shao et al. 2006). Most recently, the AurkA was also discovered to interact with H-Ras (Umstead et al. 2017).

In cells, the half-life of AurkA is two hours, and it is rapidly degraded, especially, in the beginning of the G1 phase of the cell-cycle (Honda et al. 2000). The degradation of AurkA occurs via proteasomal degradation. In order that it will become degraded, it is tagged for the degradation process by ubiquitination by the anaphase-promoting complex (APC). For instance, hEFL1 shields AurkA from this proteasomal degradation, and thereby the depletion of hEFL1 leads to AurkA degradation (Ice et al. 2013). Nonetheless, the AurkA protein levels are not only regulated by the proteasomal degradation, but also by the expression of the protein. For example, AurkA levels can be downregulated by micro-RNA, miR-124 (Qiao et al. 2017). Furthermore, transcriptional repression of AurkA has been identified to occur at least by Homeobox protein SIX3 (SIX3) (Yu et al. 2017).

At the cellular scale, the AurkA is involved in numerous critical functions, especially related to cell proliferation (Hirota et al. 2003). These functions include centrosome maturation and separation, mitotic entry, microtubule nucleation, mitotic spindle assembly, bipolar spindle microtubule formation, cytokinesis and mitosis exit (Tang et al. 2017a). In other words, AurkA is a crucial player in the G2-M progression in the cell-cycle. Therefore, unsurprisingly, AurkA knock-out mice die during embryogenesis (Lu et al. 2008).

In cancer, mutations in AurkA are rarely observed (COSMIC v.81, Forbes et al. 2017). Accordingly, the overexpression seems to be the key, and responsible for the oncogenic properties of AurkA. In fact, amplification of AURKA is observed in cancers of the large intestine (7.4%) and the stomach (5.2%) (COSMIC v.81 CONAN; Forbes et al. 2017). The overexpression is more frequent in various cancers, which is not surprising, as AURKA expression is linked to several common oncogenes; for example, in HCC, it belongs to the primary induced genes by Myc (Kress et al. 2016) and also its expression can be elevated via NRAS and CDKN2A activation (Dauch et al. 2016). AurkA is overexpressed in a wide variety of human cancers, such as colorectal cancer (Goos et al. 2013), ovarian cancer (Landen et al. 2007), esophageal squamous cell carcinoma (Tong et al. 2004), gastric carcinoma (Kamada et al. 2004) and, importantly, overexpression of AurkA is frequent in HCC and correlated with advanced disease (Jeng et al. 2004). Furthermore, especially the co-overexpression of AurkA with AurkB has been associated with a poor prognosis in HCC (Liu et al. 2017). Interestingly, it also seems that balanced AurkA levels are crucial for genomic stability, as heterozygous AURKA results in a high-rate of aneuploidy and is tumorigenic (Lu et al. 2008).

Obviously, due to its critical role in cell proliferation, AurkA has an important role in cancer, not only role in cellular proliferation, but it also seems to contribute to the other hallmarks of cancer (Hanahan and Weinberg 2011). Correspondingly, in recent years, the reports have been emerging of AurkA’s role in cell migration, adhesion, angiogenesis and invasion in different cancers. For example, migration and adhesion in ovarian cancer cells (Do et al. 2014), invasion of breast cancer cells (Xia et al. 2017), epithelial-mesenchymal transition (EMT) and metastasis in HCC (Chen et al. 2017); angiogenesis, proliferation, invasion, migration in ovarian cancer (Wang et al. 2016); migration and invasion in head and neck squamous cell carcinoma (Wu et al. 2016); migration in COS-7 cells (Mahankali et al. 2015).

To summarize, AurkA has several crucial functions in various crucial cellular processes due to its many protein–protein interactions, and it has a pivotal role in tumor growth and aggressiveness.
4.1.2 Aurora kinase A inhibitors

Numerous aurora kinase inhibitors have been developed to date. Currently, there exist data on 558 different molecules that inhibit AurkA (Hu et al. 2017). Moreover, a Pharos search resulted in activity data of 870 molecules (Nguyen et al. 2017). Because of the vast amount of AurkA inhibitors, only selected inhibitors will be discussed here in more detail. These include the compounds that have reached clinical trials, and other interesting and notable inhibitors, which display distinctive characteristics compared to regular inhibitors.

The kinase inhibitors can be classified in different subtypes. This classification is made on the basis of the kinase conformation to which the ligand preferably binds, and to the exact binding site of a ligand. The conformations are classified to the active or inactive conformation, which is reflected by the activation-loop (also known as the T-loop). In the active conformation, the T-loop is open, and in the inactive conformation it is closed to the ATP-binding cleft. In figure 11, the T-loop is in the active phosphorylated (p-T288) conformation. Furthermore, one aspect of the conformational classification is the kinase’s DFG-conformation (DFG residues: 274–276 in AurkA), stated as DFG-in or DFG-out conformations. This can also be something in-between; for instance, DFG-up conformation has been also reported. The type I inhibitors bind to the DFG-in conformation (Dar & Shokat 2011). The type II inhibitors bind to the inactive conformation to the ATP-binding pocket, where DFG residues are in the DFG-out conformation. The type I ½ inhibitors bind to the inactive conformation with DFG-in (Zuccotto et al. 2010). The type III inhibitors are allosteric inhibitors that bind next to the ATP-site. Type IV inhibitors are allosteric inhibitors, which do not bind next to the ATP site (Gavrin & Saiah 2013). Type V inhibitors are so-called bivalent inhibitors, which exhibit characteristics from different classes (Lamba & Ghosh 2012). Covalent inhibitors, which can be put into their own class, are designated as type VI inhibitors (Roskoski 2016).

However, this classification may be rather arbitrary, as in real-life, the situation is most often dynamic and not an on-off, which is usually observed in the crystallized kinase-inhibitor complexes. For instance, with the CD532 inhibitor, which is considered as a type II inhibitor, where the T-loop is in the inactive closed conformation, the activation loop is really in an equilibrium where the inactive conformation is 1.8-fold more populated compared to the inactive one (Gilburt et al. 2017). Due to this equilibrium shift with the crystal packing forces, the loop is closed in the crystal structure (PDB ID: 4J8M) (Gustafson et al. 2014). Thus, the types are not disclosed here for all of the inhibitors.

Furthermore, the selectivity with the kinase inhibitors is always a major issue, as there exist more than 500 kinases in the human kinome (Fabbro et al. 2015). Moreover, due to the high-similarity of the kinase domain within the aurora kinase family, the majority of the developed AurkA inhibitors hit all the three kinases in the aurora kinase protein family, and thus are designated as pan-selective or pan-aurora inhibitors (Fig. 12). These include ABT-348, AMG 900, AT-9283, CYC116, PF-03814735, PHA-739358 (danusertib), R763/AS703569 (cenicertib), SNS-314 and VX-680 (MK-0457, tozasertib).
ABT-348 (ilorasertib) is a multi-target kinase inhibitor discovered by Abbott (Curtin et al. 2012). In addition to aurora kinases, ABT-348 inhibits VEGFR, PDGFR and Src families (Glaser et al. 2012). In fact, ABT-348 inhibits more potently AurkB and AurkC (IC<sub>50</sub>-values 7 nM and 13 nM, respectively) than AurkA (IC<sub>50</sub> = 189 nM). Moreover, it inhibits VEGFR1-3, FLT-3, CSF-1R at single-digit nanomolar levels, and c-KIT and PDGFRα-β in 45 nM, 16 nM and 11 nM, respectively. ABT-348 displayed an acceptable safety profile in a phase I study in patients with hematological malignancies (Garcia-Manero et al. 2015). It is currently in phase II trials in patients with CDKN2A deficient solid tumors (NCT02478320).

The aurora kinase inhibitor, AMG 900, developed by Amgen, inhibits AurkA and AurkB at similar single-digit nanomolar concentrations (IC<sub>50</sub> values of 7 nM and 8 nM, respectively) (Geuns-Meyer et al. 2015). It inhibits p38 with IC<sub>50</sub> value of 53 nM and TYK2 with an IC<sub>50</sub> value of 220 nM (Payton et al. 2010). Moreover, it displays low-nanomolar K<sub>d</sub> values for several off-
targets: DDR1 2nM; LTK 13 nM, DDR2, 31 nM; FES 61 nM; and Tie-1 97 nM (Payton et al. 2010). It has been in phase I clinical trials for acute myeloid leukemia (Kantarjian et al. 2017).

A fragment-based drug discovery program conducted by Astex Therapeutics resulted in the emergence of AT-9283 (Howard et al. 2009). It has an IC50 value of 3 nM for both AurkA and AurkB. AT-9283 is also high-potent (IC50 value < 4 nM) JAK2, JAK3 and Abl T315I inhibitor. Overall, it is a rather promiscuous kinase inhibitor, as it inhibits nearly 50 different kinases with IC50 values of less than 100 nM. The crystal structure of the AT-9283 with AurkA was published (PDB ID: 2W1G), confirming it as a type I inhibitor. It was tested in phase I clinical trials in patients with advanced solid tumors, also in children, and in patients with haematological malignancies (Arkenau et al. 2012, Dent et al. 2013, Foran et al. 2014, Moreno et al. 2015). However, in a phase I/II trial of haematological cancers in children, it failed to show efficacy (Vormoer et al. 2017).

Cyclacel reported an inhibitor CYC116, with Ki values of 8 nM and 9.2 nM for AurkA and AurkB, respectively (Wang et al. 2010). Moreover, CYC116 also inhibits CDK2 (K: 390 nM), FLT3 (K: 44 nM), CDK9 (K: 480 nM), ribosomal protein S6 kinase (p70S6K) (K: 540 nM), SRC (K: 820 nM) and VEGFR2 (K: 44 nM). It was in phase I clinical trials in patients with advanced solid tumors; however, no results have been reported (NCT00560716).

PF-03814735 is a dual AurkA and AurkB inhibitor (Jani et al. 2010). It is somewhat more potent towards AurkB with the IC50 value of 0.8 nM, whereas it inhibits AurkA with an IC50 value of 5 nM. Interestingly, the PF-03814735 displayed specific efficacy in Myc-driven small cell lung cancer models (Hook et al. 2012). The inhibitor was in phase I study in patients with advanced solid tumors, displaying tolerable toxicity profile, but its efficacy was limited (Schöffski et al. 2011). Further clinical trials with PF-03814735 were not conducted.

PHA-739358 (danusertib) inhibits aurora kinases with IC50 values 13 nM for AurkA, 79 nM for AurkB and 61 nM for AurkC (Fancelli et al. 2006, Carpinelli et al. 2007). Danusertib has been crystallized with AurkA (PDB ID: 2J50) (Fancelli et al. 2006). Moreover, it has been reported to inhibit ABL, TRKA, RET, FGFR1 (with IC50 < 50 nM), and LCK, VEGFR3, C-KIT, VEGFR2, CDK2/cyA, STLK2, FLT3 (with IC50 < 700 nM) (Fancelli et al. 2006). Additionally, it was disclosed to inhibit receptor tyrosine kinase EPHA2 (Ephrin type-A receptor 2) with an IC50 of 4.4 nM, and the crystal structure of the complex was published (PDB ID: 519Z) (Heinzleir et al. 2016). PHA-739358 in a complex with Mammalian Sterile20-like kinase 3 (MST3) was published (PDB ID: 4Q09) (Olesen et al. 2016), and with Tyrosine-protein kinase ABL1 (ABL) mutant T315I (PDB ID: 2V7A) (Modugno et al. 2007). Danusertib has been investigated in several clinical trials, in phase I for hematological malignancies (Borthakur et al. 2015 and for advanced or metastatic solid tumors (Steeghs et al. 2009); in phase II for castration-resistant prostate cancer (Meulenbeld et al. 2013) and advanced or metastatic solid tumors of various tissues of origin (Schöffski et al. 2015). R763/AS703569 (also known as MSC1992371A; cenicertib) was developed by Rigel Pharmaceuticals via a phenotypic screen (McLaughlin et al. 2010). It inhibits AurkA, AurkB and AurkC with IC50-values of 4 nM, 4.8 nM and 6.8 nM, respectively. Moreover, it displays off-target activities: at 100 nM it inhibits (> 70% inhibition) ABL1, CHK1, FGFR3, FLT1, FLT3 and VEGFR2. It underwent phase I clinical trials for haematological malignancies and for solid tumors (Graux et al. 2013, Mita et al. 2014); thereafter, it was discontinued.

SNS-314 (Oslob et al. 2008), developed by Sunesis Pharmaceuticals, was described in a complex with a mutant AurkA (PDB ID: 3D15). It is a pan aurora inhibitor; it inhibits AurkA with the IC50 value of 9 nM, AurkB with 31 nM and AurkC with 3 nM (Arbitrario et al. 2010). Moreover, its off-targets include (with the respective IC50 values): TrkA, 12 nM; TrkB, 5 nM; Flt4, 14 nM; Fms, 15 nM; DDR2, 82 nM; Axl, 84 nM; c-Raf, 100 nM. SNS-314 displayed YAP mediated apoptosis in a HCC cell-line, and outperformed MLN8237 in HepG2 cell viability assay (Liu et al. 2017). It has undergone a phase I clinical trials for advanced solid tumors (NCT00519662).

VX-680 (MK-0457, tozasertib), developed by Vertex Pharmaceuticals, is a type II inhibitor and inhibits aurora kinases with IC50 values of 0.7 nM (AurkA), 18 nM (AurkB) and 4.6 nM (AurkC) (Harrington et al. 2004). It also binds to the AurkA-TPX2 complex (PDB ID: 3E5A) (Zhao et al.
In addition to Aurora kinases, VX-680 inhibits other kinases, including Proto-oncogene tyrosine-protein kinase Src (SRC), FMS-like tyrosine kinase 3 (FLT3), Tyrosine-protein kinase Fyn (FYN), Tyrosine-protein kinase ITK/STK (ITK) and Tyrosine-protein kinase Lck (LCK) (Harrington et al. 2004). Moreover, VX-680 binds to ABL1 H396P (PDB ID: 2F4J) (Young et al. 2006) and ABL2 (PDB ID: 2XYN) (Salah et al. 2011). It has been examined in phase II clinical trials of haematological malignancies where it displayed low efficacy with adverse effects (Seymour et al. 2014). Moreover, it underwent a phase IIa trial for the lung cancer (NCT00290550).

Not only pan-selective inhibitors have been developed, but also selectivity for AurkA among the other aurora kinases has been pursued. The following inhibitors are examples of the inhibitors that have been stated as selective for AurkA: Aurora A inhibitor I (S1451), ENMD-2076, MLN8054, MLN8237 (alisertib) and VX-689 (MK5108) (Fig. 13).

Genentech reported the Aurora A inhibitor I (S1451) with a thousand-fold selectivity for AurkA over AurkB (Aliagas-Martin et al. 2009). In another study, it displayed IC50 values of 3.2 nM, 1,380 nM and 432 nM for AurkA, AurkB and AurkC, respectively (Kwiatkowski et al. 2012). It was shown to be effective in a CML cell line (Yuan et al. 2012). However, in our study, we observed that S1451 was ineffective in various HCC cell-lines (Dauch et al. 2016).

CASI pharmaceuticals compound ENMD-2076 is a multikinase inhibitor, which also targets FLT3, SRC, c-KIT, FGFR1, VEGFR2 (Tentler et al. 2010). Moreover, it also inhibits AurkB with an IC50 value of 350 nM (recombinant protein), showing 25-fold more potent IC50 value for AurkA; however, this IC50 values is only 3.5-fold more potent for AurkA in a cellular setting. It has been in phase I clinical trials for haematological malignancies (Yee et al. 2016). It is currently in phase II trials for fibrolamellar carcinomas (NCT02234986).

MLN8054 (PDB ID: 2WTV), developed by Millennium Pharmaceuticals, inhibits AurkA with the IC50-value of 4 nM (Manfredi et al. 2007). It was reported to have 43-fold selectivity for AurkA

![Figure 13. Selective AurkA inhibitors.](image-url)
over AurkB (IC\textsubscript{50} = 172 nM). However, in another assay, the selectivity was observed to be 9.3-fold, with IC\textsubscript{50}-values of 10 nM and 93 nM for AurkA and AurkB, respectively (Aliagas-Martin et al. 2009). It displays some inhibition of other kinases, but not at the level of Aurora kinase inhibition. These inhibition levels are LCK (IC\textsubscript{50}-value 3.2 \(\mu\)M), PKA (IC\textsubscript{50}-value 19 M), CKII (IC\textsubscript{50}-value 20.5 \(\mu\)M), CHK2 (IC\textsubscript{50}-value 28 \(\mu\)M), and MLN8054 displays inhibition of Abl, EphA1 and Yes (> 70% inhibition at 1 \(\mu\)M). In humans, some off-target related somnolence was reported as a dose limiting toxicity (Sells et al. 2015). After binding to AurkA, the MLN8054 shifts the equilibrium of the activation loop towards the closed state (Gilburt et al. 2017). However, when TPX2 is also bound, MLN8054 does not seem to have any effect to the activation loop dynamics. It underwent phase I clinical trials for advanced malignancies and for advanced solid tumors (NCT00652158, NCT00249301).

MLN8237 (alisertib), a derivative of MLN8054, inhibits AurkA with an IC\textsubscript{50}-value of 1 nM (Sells et al. 2015). MLN8237 was recently identified to bind also to p38 mitogen-activated protein kinase (p38) and laminin receptor (LAMR) with DNA-based affinity labeling (Wang et al. 2017). Alisertib has been the most widely studied AurkA inhibitor in clinical trials to date. Several clinical trials have been conducted, including a phase I/II study for castration-resistant prostate cancer progressing on abiraterone, displaying no efficacy (Lin et al. 2016); a phase I trial of myeloid leukemia (Fathi et al. 2017); a phase I trial of solid tumors with docetaxel (Graff et al. 2016); phase II trials have been conducted in hematological malignancies (Cohen et al. 2017); advanced or metastatic sarcoma, displaying some progression-free survival (Dickson et al. 2016); leiomyosarcoma, with lack of efficacy (Hyman et al. 2017). Moreover, it is currently in a phase III trial for T-Cell lymphoma (NCT01482962). In our study, the MLN8237 displayed efficacy in the specific TP53 deficient dependent HCC in vitro and in vivo; putatively via the distortion of AurkA–Myc interaction (Dauch et al. 2016). Conversely, the compound displayed only a modest effect in the N-Myc–AurkA interaction in neuroblastoma tumor cell-lines (Gustafson et al. 2014), further highlighting the existence of potential differences in AurkA-binding of these transcription factors. Moreover, in small cell lung cancer cell-lines, it displayed no effect on Myc protein levels (Mollaoglu et al. 2017).

VX-689 (MK-5108) is stated as a selective AurkA inhibitor, as it inhibits AurkA with IC\textsubscript{50} of sub-nanomolar level (Shimomura et al. 2010). Moreover, the K\textsubscript{i} was reported to be below 10 pM (de Groot et al. 2015) However, it still inhibits AurkB and AurkC with an IC\textsubscript{50} value of ~10 nM. Moreover, it has been in phase I clinical trials for advanced solid tumors, and it is unclear whether or not it inhibits AurkB in vivo (Amin et al. 2016). The structure of VX-689 has been published in complex with AurkA (PDB ID: 5EW9) (de Groot et al. 2015), it binds to a tightly folded conformation next to the hinge-region.

Overall, as is quite evident from above, although these inhibitors were classified as selective inhibitors, the selectivity is a rather arbitrary term within the class for some compounds, and caution should be made when interpreting the results of these selective inhibitors. In fact, some of the inhibitors, such as MLN8054 and MLN8237, will even display a diminished selectivity window, when TPX2 is bound to AurkA (de Groot et al. 2015).

Finally, a couple inhibitors that could not be included in the pan-aurora or in the selective inhibitors need to be mentioned here. First, CD532 (Fig. 14) which inhibits AurkA with IC\textsubscript{50} of 48 nM and the crystal structure of the complex has been published (PDB ID: 4JBM) (Gustafson et al. 2014). Moreover, it induces a conformational shift in AurkA, pushing the G-loop and the Ca-helix, which prevents N-Myc or Myc binding (Gustafson et al. 2014, Dauch et al. 2016, Richards et al. 2016). CD532 shifts the activation loop equilibrium towards the closed state (Gilburt et al. 2017). Interestingly, the off-target effects of CD532 i.e. binding to other kinases (even the aurora kinases) or proteins have not been disclosed to date. Therefore, it cannot be classified in either of the formerly mentioned groups. Furthermore, the other inhibitor worth mentioning here, is a type IV inhibitor. In fact, AurkinA is the only allosteric inhibitor of AurkA that has been reported to date (Janeček et al. 2016). AurkinA binds to the TPX2 binding site on top of the N-lobe and
inhibits to some extent its kinase activity. However, the compound is not very potent, and the \( K_d \) value was reported to be only 3.77 \( \mu M \).

![Figure 14. Structure of CD532.](image)

Most of the clinical trials for AurkA inhibitors have been conducted against hematological malignancies. However, some of them have been also tested against solid tumors, where they have displayed limited efficacy. This emphasizes the fact that plainly the kinase inhibition is not enough for the efficacy with AurkA inhibitors in solid tumors. In the clinical trials, some toxicity with the AurkA inhibition has been observed. Neutropenia has been linked as a dose-dependent primary toxicity issue to appear with Aurora kinase inhibition. For instance, the neutropenia with MLN8237 (alisertib) was observed in 50\% of the patients (Falchook et al 2016). Moreover, leukopenia and thrombocytopenia were observed in 38\% and 21\% of the patients, respectively. Overall, the inhibitors seem to be generally well-tolerated.
4.1.3 The efficacy problem with Aurora kinase A inhibition

Due to the vast number of kinases and their similarity, the extent of the selectivity is also clearly still an issue with AurkA inhibitors. In fact, very few of the kinase inhibitors developed to date are truly selective (Anastassiadis et al. 2011, Klaeger et al. 2017). There is no exact data available detailing which of the off-targets can be tolerated with the treatment of solid tumors i.e. which off-targets would be clearly diminishing the efficacy, or are there any that could even improve the efficacy. This off-target tolerability may differ among malignancies from distinct tissues. Moreover, toxicity related issues may also appear via the off-targets.

To improve the inhibitor selectivity among the kinases, Badrinaryan and Sastry published a study where they tried to identify specificity hotspots for Aurora kinase inhibitor design (Badrinarayan & Sastry 2014). They pointed out that residues P144, Q177, Q185, L210, P214, L215, T217, R220, W277, V279 and H280 mainly differ with the sequentially similar kinases within the active site. Therefore, these may be the most important residues to focus with direct interactions to achieve selectivity between similar kinases. Furthermore, the sub-selectivity among the aurora kinases is also crucial. Due to the high-similarity within the ATP-binding cleft among the aurora kinases, the targeting of AurkA specific residues L215, T217 and R220 was suggested (Sarvagalla and Coumar 2015). The corresponding residues in AurkB and AurkC are R215, E217 and K220. The residue L/R215 side-chain is oriented towards the solution; thus; making it unreachable. The residue T/E217 is located in the solution interface, but not easily accessible. The residue R/K220 which is further away in the solution and they contain similar characteristics. Resulting in an almost impossible scenario to the selectivity in the direct interactions in the amino-acid level. Therefore, the dynamical differences among the Aurora kinases need to be identified; these could be utilized in defining selectivity ranges among the family. As even though the kinase domains of aurora kinases appear to exist in a similar conformation, it does not imply that they are equal and represent similar dynamics among themselves (Tzeng and Kalodimos 2012).

Some of the AurkA inhibitors have been stated to be AurkA selective (see section 4.1.2). However, the observed residence times of the inhibitors compromise their stated selectivity in vivo. For long, the only reported residence times for AurkA inhibitors have been those reported for the fluorescent probe derivatives of MLN8237 and VX-689 i.e. 500–3000 s, and the presence of TPX2 increased the residence time (Lavogina et al. 2014). Recently, the residence times for AMG900, danusertib, GSK1070916 (MLN5108) and MLN8054 were reported (Willemsen-Seegers et al. 2017). These inhibitors exhibit clearly shorter residence times with AurkA compared to AurkB. Moreover, the ‘selective’ AurkA inhibitor MLN8054 (reported as 43-fold selective for AurkA), displayed a really poor residence time for AurkA (76 s), whereas the residence time for AurkB was 5820 s. Danusertib displayed residence times of 1153 s for AurkA and 15976 s for AurkB, whereas tozasertib displayed 566 s and 11473 s for AurkA and AurkB, respectively. These residence time differences with aurora kinases imply that even though selectivity for AurkA seems to exist based on the IC50 values, the difference in the residence time could actually alter dramatically the selectivity in vivo.

Probably the most crucial efficacy related issue with AurkA inhibition is related to its non-kinase-related role in the cellular signaling. At present, most of the focus in AurkA inhibitor development has been placed on achieving the best kinase activity inhibition, and selectivity over other Aurora kinases. However, we observed that at least in a particular subtype of HCC, the ability of the inhibitors to distort non-kinase activity interactions of AurkA (in addition to the kinase inhibition), determined the efficacy of the compounds (Dauch et al. 2016). We discovered that only the inhibitors that would also induce a conformational change in AurkA were effective. This conformational change distorts the AurkA–Myc interaction, and in that way leads to Myc’s proteasomal degradation (Fig. 15a). We created a putative model of the AurkA–Myc interaction, which provided a putative explanation for the observed differences between the inhibitors (Fig. 15b). In the model building, we first mapped the phosphate binding hot-spots in AurkA’s structures and, interestingly, discovered that the most energetically favorable hot-spot is not present in the conformationally changed CD532–AurkA complex. Next, we used this phosphate
hot-spot as an anchor point for the docking of the phosphorylated Myc peptide, and selected the best complex based on the interactions as for the putative model.

Figure 15. AurkA–Myc interaction is distorted only by the conformational change inducing inhibitors. (a) The phosphorylated Myc binds AurkA without overlapping the ATP-binding site; therefore, the traditional AurkA inhibitors have no effect on Myc binding, whereas the conformational change inducing inhibitors prevent MYC binding. (b) The putative structure of AurkA–Myc complex. Danusertib (PDB ID: 2J50) fits nicely in its binding site when Myc is bound, whereas CD532 (PDB ID: 4JM8) does not fit into this conformation. The binding pocket illustrated with wheat mesh surface. Myc residues. AurkA as gray cartoon. Myc (residues 55–67) displayed as cyan stick model with the phospho-groups as ball and stick in the phosphorylated T58 and S62 residues.

Our findings suggest, at least in a specific cellular setting, merely exhibiting the kinase activity inhibition of AurkA is not sufficient to achieve a therapeutical response (Dauch et al. 2016). We discovered that AurkA can shield Myc from degradation by the ubiquitin ligases. Moreover, only
the inhibitors inducing a conformational change in AurkA are able to prevent the Myc binding to AurkA, and thereby prevent AurkA mediated shielding of the Myc from the degradation. The exact binding of Myc to AurkA is unknown to date, and there is no structural data available. The N-Myc–AurkA interaction has been disclosed via their X-ray structure (Richards et al. 2016). However, unlike the Myc-AurkA interaction, the N-Myc–AurkA interaction is not dependent on N-Myc phosphorylation. Correspondingly, in the complex with AurkA, the N-Myc is unphosphorylated (PDB ID: 5G1X). K-Ras is linked to Myc phosphorylation, as ERK (downstream from Ras in MAPK-pathway) phosphorylates S62 of Myc, which leads to the subsequent phosphorylation at T58 by GSK3β (Farrell and Sears 2014). Moreover, AURKA belongs to the primary induced genes by Myc in HCC (Kress et al. 2016). Therefore, the active Myc protein produces more shielding AurkA for itself.

As clearly noted above (see section 4.1.1), AurkA has numerous protein–protein interactions. Therefore, one important question is, does there exist another non-kinase activity related to the specific interaction and is it its distortion which is also crucial for a drug’s efficacy. Moreover, the specific cellular setting may define this; for example, the AurkA – Myc interaction is only crucial in a specific subtype of HCC. Importantly, this exemplifies that accurate patient selection for the clinical trials is needed, as the efficacy may be both specific and dependent on the tumor’s genetic background.

To date, the drug discovery for AurkA has focused only on inhibiting its kinase activity. At present, we and others have disclosed crucial other interactions of AurkA with other proteins that are not dependent on its kinase activity, and they seem to be really crucial in the efficacy perspective. Therefore, in order to achieve efficacy with the AurkA inhibition, not only does the molecule need to be selective, but it also must have an effect on non-kinase activity related protein–protein interactions.
4.2 AUTOTAXIN

4.2.1 Autotaxin as drug target

The ENPP2 gene product, Autotaxin (ATX) (EC number: 3.1.4.39), is the only member of the ectonucleotide pyrophosphatase/phosphodiesterase family (ENPP) that has lysophospholipase D (lysoPLD) activity (Umezu-Goto 2002). The secreted enzyme, ATX, hydrolyses lysophosphatidylcholine (LPC) to produce lysophosphatidic acid (LPA) and choline (Scheme 1). The produced LPA is a bioactive signaling lipid, which binds and signals mainly through specific G-protein- coupled lysophosphatic acid receptors 1 to 6 (LPA1-6) (Kihara et al. 2014). ATX is ubiquitously expressed in humans and is present in blood (Aoki et al. 2002). In order to result in specific spatiotemporal signaling, ATX is recruited to the cell surface by activated β3-integrins (Pamuklar et al. 2009, Leblanc 2014). ATX binds to activated integrins via its somatomedin B-like (SMB) domains (Hausmann et al. 2011), and in that way becomes tethered to plasma membrane, promoting the allocated LPA-signaling. This results in a localized increase in the LPA concentration and a specific biological response, as LPA stimulates cell migration, proliferation and survival (Moolenaar et al. 2004), induces platelet aggregation and chemotaxis (Jalink et al. 1993), smooth muscle contraction (Tokumura et al. 1994), neurite remodeling (Fukushima et al. 2002) and ion channel activity (Itifinca et al. 2007). Consequently, ATX contributes to various physiological and pathophysiological processes, such as embryonic development (Moolenaar et al. 2013), wound healing (Lee et al. 2013), inflammation (Knowlden and Georas 2014), vascular (van Meetereen et al. 2006) and neural development (Fotopoulou et al. 2010), and in tumor growth, metastasis (Leblanc and Peyruchaud 2015) and chemoresistance (Brindley et al. 2013).

Scheme 1. LPC is hydrolyzed by ATX resulting in LPA and choline.

Accordingly, ATX is a potential therapeutic target in various diseases. In cancer, the ATX–LPA signaling axis plays a role in various tumor types (Liu et al. 2009, Houben and Moolenaar 2011). Genetic silencing of ATX and LPA receptors expression in mouse models have revealed the importance of ATX–LPA axis in cancer development (Liu et al. 2009). Moreover, increased expression of ATX was linked with chronic liver disorders and hepatocellular carcinoma (Kaffe et al. 2017, Yamakazi et al. 2017), thus exemplifying its role in HCC, and further in HCC development (Lopane et al. 2017). Interestingly, the other LPA producing enzyme, phosphatidic acid-selective phospholipase A1α (PA-PLA1α), displays similar levels in HCC as in healthy liver tissue (Enooku et al. 2016). This further points to a crucial role of the ATX, and not the LPA alone, in the liver tumors. Moreover, ATX–LPA signaling has been linked to cholestatic diseases and pruritus (Beuers et al. 2014), and in fact, it was later disclosed that the bile-salts are able to inhibit ATX (Keune et al. 2016). It remains to be seen if these bile-salts play a role in HCC or its development. Correspondingly, the role of ATX has been suggested as being important in pancreatic cancer (Kadekar et al. 2012, Quan et al. 2017). Moreover, the ENPP2 is amplified in tumors of the breast (8%), the ovary (7%), and the liver (5%) (COSMIC v80, Forbes et al. 2017). Thus, highlighting its role in a wide spectrum of solid tumors.

The ATX encoding gene ENPP2 is located in chromosome 8, and it is the only member of the ectonucleotide pyrophosphatase/phosphodiesterase family (ENPP) that has lysophospholipase D (lysoPLD) activity (Umezu-Goto et al. 2002). Three ATX isoforms exist via the alternative splicing: ATXα (105kDa; 915 residues), ATXβ (99kDa; 863 residues) and ATXγ (102kDa; 888 residues). The
isoforms’ amino acid composition is nearly identical, as the ATXα residues 324–376 are absent from ATXβ and ATXγ, and the ATXγ has an extra 25 residues insert in the corresponding ATXα position 645–646. As a glycoprotein, ATX has four glycosylation sites: N54, N411, N525 and N807. ATX is expressed in all tissues (Uhlén et al. 2015), and its expression levels are regulated by the RNA binding proteins Human antigen R (HuR) and ARE/poly(U)-binding/degradation factor 1 (AUF1) (Sun et al. 2016). HuR enhances ATX mRNA stability, whereas AUF1 promotes its degradation.

The structure of ATX consists of four domains: two somatomedin B (SMB)-like domains SMB1 and SMB2, catalytic phosphodiesterase (PDE) domain and a nuclease (NUC)-like domain (Fig. 16a). The SMB1 (55–98) and SMB2 (99–143) are responsible for the integrin binding, and tethering ATX to the plasma membrane (Hausmann et al. 2011), which is indeed crucial for the cell migration (Wu et al. 2014). The catalytic site, located in the PDE domain, contains two Zn2+-ions right next to the catalytic threonine, T210 (T209 in mouse). Moreover, the substrate binding site contains a lipophilic cavity (Fig. 16b), which can be occupied by the LPC’s hydrophobic tail during the enzymatic reaction. Furthermore, it has been suggested that the product, LPA, leaves the site through a tunnel (also known as channel) which is linked to the cavity (Fig. 16c) (Hausmann et al. 2011, Nishimasu et al. 2011). The NUC domain interplays with the catalytic domain, and is important for the catalytic activity, secretion and stability of ATX (Jansen et al. 2009, Nishimasu et al. 2012).
Figure 16. Structure of Autotaxin. The structure of ATX consist of four domains: SMB1 and SMB2; phosphodiesterase (PDE) domain, grey; nuclease domain, red (a). The binding site of the substrate is located in the PDE domain, which contains two Zn^{2+}-ions (blue, spheres) and the catalytic Thr210 (sticks) in the active site. The LPC/LPA lipophilic tail binds to a lipophilic cavity (highlighted with green) (b). There exists also a tunnel in the pocket (highlighted with cyan), which is the putative exit tunnel for the product, LPA, from the site (c). PDB ID: 5KXA (Bain et al. 2017).

Since it is an extracellular secreted enzyme, ATX represents a highly attractive drug target in cancer. In addition to its implied role in HCC and PDAC, it has been shown that pharmacological inhibition of ATX is well tolerated, at least in adult mice (Katsifa et al. 2015). Therefore, clear potential exists to design novel small-molecule inhibitors targeting this enzyme.
4.2.2 Autotaxin inhibitors

As the overall data of Autotaxin is constantly increasing, reaching over 100 publications alone in 2016, similar to the data of ATX inhibition. To date, almost 1,400 ATX associated bioactivities are available in ChEMBL (Bento et al. 2014). This section covers briefly the developed ATX inhibitors, focusing on the most recent findings in ATX inhibition.

From the historical perspective, the developed ATX inhibitors can be classified in three different generation classes. The generation classification provides a good insight into the evolution of the ATX drug discovery field. The first-generation compounds include the substrate mimicking lipid-like structures, for example, S32826 (Fig. 17), which has a similar phosphate group as in LPA, and a long lipophilic tail (Ferry et al. 2008). The second-generation inhibitors consist of non-lipid-like structures, which also occupy the substrate binding site. These compounds are more drug-like, and the Zn$^{2+}$-ion binding phosphate group is replaced; for instance, by a boronic acid in HA155 (Fig. 17), and with a 2-oxo-2,3-dihydro-1,3-benzoxazole in PF8380 (Fig. 17) (Albers et al. 2010, Gierse et al. 2010). In general, most of the reported ATX inhibitors, which bind to the active site, display a similar chemotype, consisting of three elements: an acidic moiety, a spacer core and a lipophilic tail (Castagna et al. 2016). Finally, the third-generation compounds occupy a location other than the active site; these include compounds such as the (S)-25 developed by us, GLPG1690 and 3b (Fig. 17) (Desroy 2016, Banerjee et al. 2017, Pantsar et al. 2017a).

![Figure 17. Examples of Autotaxin inhibitors. S32826 (Ferry et al. 2008); HA155 (Albers et al. 2010); PF8380 (Gierse et al. 2010); GLPG1690, Galapagos NV (Desroy 2016); (S)-25 (Pantsar et al. 2017a); 3b (Banerjee et al. 2017).](image)

To date, the trend in ATX inhibitor discovery has been clearly shifted towards the third-generation compounds, which actually display significant diversification in their actual binding sites. Therefore, an improved, more up-to-date, inhibitor classification was introduced recently by Joncour et al. (Joncour et al. 2017). They produced a classification of ATX inhibitors into four classes, type I–IV, based on their binding site. The type I inhibitors occupy the catalytic site and
the hydrophobic pocket; type II inhibitors occupy only the lipophilic pocket; type III inhibitors bind to the tunnel region, also known as hydrophobic channel; and type IV inhibitors occupy the tunnel and the lipophilic pocket.

HA155, which is one of the most widely used reference compounds for ATX inhibition, is a type I inhibitor (PDB ID: 2XRG) (Hausmann et al. 2011). Furthermore, other boronic acid derivatives have been developed (Kawaguchi et al. 2013), and a few of them have been crystallized in complex with ATX (PDB ID: 3WAW, 3WAX, 3WAY).

The other commonly used ATX tool compound, which is also a type I inhibitor, is PF8380 (Gierse et al. 2010). Instead of a boronic acid, PF8380 utilizes a benzoxyal to bind the Zn\(^{2+}\)-ions in the active site. In fact, it was the first reported ATX inhibitor to represent decreased LPA levels \textit{in vivo}. Moreover, a phospho-mimic of PF8380 was recently introduced (Balupuri et al. 2017). Recently, a nanomolar inhibitor derivatized from the PF8380 was introduced. (Kutruff et al. 2017). The inhibitor is also freely available via the Boehringer Ingelheim opnMe portal (www.opnme.com).

From a lead molecule (PDB ID: 5L0B), utilizing structure-based design, Jones et al. optimization resulted in an ATX inhibitor with plasma IC\(_{50}\) value of 2 nM (Jones et al. 2016). The best compound was a triazole derivative, where the triazole targets the active site. Therefore, the compound can be considered as a type I inhibitor.

Our own drug discovery campaign against ATX, starting from a virtual screen of 2.7 million compounds (>11.8 million screened structures), resulted in the compound (S)-25 (Fig. 17) (Pantsar et al. 2017a). The inhibitor displayed IC\(_{50}\)-value of 134 nM (hATX), and was shown to inhibit LPC mediated cell-migration. The molecular modeling results combined with the experimental data suggest that (S)-25 binds to the lipophilic pocket region. In addition, our WaterMap calculations pointed to two locations for the high-energy dewetted cavity regions that would be occupied by the compound. Moreover, the binding mode is supported by the activity data, as our compound had only a minor effect to the artificial substrate pNP-TMP hydrolysis that binds to the active site and not to the lipophilic pocket, and the other isoform of the compound (R)-25 appears totally inactive, which was unable to obtain a proper docking pose to the site where (S)-25 poses converged. Therefore, (S)-25 can be stated as a type II inhibitor.

A high-throughput-screening (HTS) campaign conducted by Shah et al. resulted in the discovery of inhibitors with an imidazopyridine scaffold; the best of those, CRT0273750, displayed an IC\(_{50}\) value for ATX inhibition of 14 nM in human plasma (Shah et al. 2016). The crystal structure of the complex was published (PDB ID: 5LIA) and, interestingly, it seems to have a similar binding mode as we suggested for (S)-25.

The PharmAkea reported a PAT-series of compounds, with distinct binding modes (PDB IDs: 4ZG6, 4ZG7, 4ZG9, 4ZGA) (Stein et al. 2015). The best compound of the PAT-series; PAT-505 was shown to be effective in a mouse liver fibrosis model (Bain et al. 2017). It displayed an IC\(_{50}\)-value of 9.2 nM in human blood (20:4 LPA); in the same assay PF8380 and GLPG1690 had values of 280 nM and 82 nM, respectively. PAT-505 is believed to bind to the hydrophobic tunnel (PDB ID: 5KXA). In a screen of 55 off-target proteins, it exhibited minor activity against adenosine A3 receptor, MT1 melatonin receptor, prostaglandin E2 EP4 receptor, 5-HT5a serotonin receptor and GABA-gated Cl\^- channel.

The Galapagos ATX inhibitor GLPG1690 (Fig. 17) is the only ATX inhibitor that has reached clinical trials, and it has received designated orphan status in EU and the US for the treatment of idiopathic pulmonary fibrosis (EMA 2016, FDA 2017). GLPG1690 has been shown to bind to the hydrophobic channel and the hydrophobic regions (PDB ID: 5MHP) (Desroy et al. 2017). Interestingly, the most promising ATX inhibitors to date, GLPG1690 and PAT-505, are both type IV inhibitors, utilizing the tunnel and the lipophilic pocket.

The binding mode has not been disclosed for some inhibitors; for instance, binding data have not been disclosed for the patented compounds from Ono Pharmaceutical Co. Ltd. (Ohata et al. 2012).
Furthermore, bile salts have been found to bind to the tunnel region and inhibit ATX (Keune et al. 2016). It remains to be seen if there is a link between the bile-salt mediated ATX inhibition and liver disease development, and further to HCC. Consequently, Keune et al. utilized these bile-salt as lead compounds and made derivatives, which extended the binding to the lipophilic pocket (Keune et al. 2017). Of these, the best derivative displayed a Ki value of 6 nM, and in vivo in mice its administration halved LPA levels.

In addition, an aptamer for ATX has been introduced (Kato et al. 2016). The aptamer binds to the vicinity of the active site (PDB ID: 5HRT), putatively blocking the LPC binding. The best aptamer inhibits LPA production in human serum nearly fully at 0.1 µM. Moreover, it displayed efficacy in vivo in a pulmonary fibrosis mouse model.

Finally, first metal-based ATX inhibitors have been introduced (Kang et al. 2017). The best inhibitor was a rhodium(III) based complex, which was shown to bind directly to ATX in a thermal shift assay, and also inside the cells with a cellular thermal shift assay (CETSA). However, the binding site was not disclosed. Overall, these compounds seem to display rather low activity (IC50 value of 2.10 µM with an unnatural ATX substrate).

### 4.2.3 The efficacy problem with Autotaxin inhibition

The ATX inhibition has produced promising results to date. However, a few issues regarding to efficacy problem with ATX inhibition/inhibitors have emerged, and this section is devoted to these issues.

First, one crucial aspect of suitable efficacy achievement is traditional, does the compound have appropriate pharmacokinetics. Since the ATX is extracellularly located in the circulation, the high-lipophilicity of the compounds may be an issue, as the drugs with this characteristic usually have a short half-life in circulation (Oie 1986). One tactic that could be employed here, would be to optimize the albumin binding of the drug candidate since optimal plasma protein binding properties may enhance the half-life of the drug in circulation (Sleep et al. 2013). Obviously, the binding with albumin should not be too extensive, so that the free-drug will be still available for ATX. Most of the developed ATX-inhibitors to date, tend to exhibit lipophilic characteristics (Castagna et al. 2016). Actually, this is not surprising as the natural ligand LPC's tail has high-lipophilic characteristics, thereby the binding cavity also displays this characteristic. For instance, the 2,4-dihydropyran[2,3-c]pyrazole series developed by us (Pantsar et al. 2017a) exhibited generally a clear trend in correlation with higher-lipophilicity and ATX inhibition activity. This highlights that the ligand-lipophilicity efficiency (LLE) should receive a major emphasis in any ATX drug discovery programme (Leeson & Springthorpe 2007).

As mentioned above, there seems to be a high-risk with the cavity binding small-molecules to obtain undesired pharmacokinetic profile, leading to inefficacy. Consequently, inhibitors (LM350) with a less lipid-like scaffold, have been lately introduced (PDB ID: 5LQQ) (Miller et al. 2017).

When compared to kinase inhibitors (see section 4.1.2), where at least off-target kinase activity characterization is a standard routine (partly due to high-similarity and the quantity of the kinases), this is not usually the case for the off-target monitoring with ATX inhibitors. There seems to be a trend that with fresher – not-so-much studied – drug targets, this is quite common. However, this would be beneficial not only for the project at hand, but also for all the future ATX drug discovery projects aimed at identifying as many ATX inhibitor related off-targets as possible. The more caveats that are known regarding the most common ATX off-targets, the lower the chance for similar failures.

One important aspect of efficacy, the drug-target residence time (Copeland 2016), has not been covered at all in the literature with regard to ATX-inhibition. As the residence times of the inhibitors have not been disclosed, not much can be speculated with regard to how this parameter will influence the efficacy. This important question should be answered by future research.

Another aspect in the efficacy problem with ATX inhibitors, is the widely used unnatural substrates in the inhibition assays. The most commonly used unnatural substrates include pNP-
TMP, bis-pNPP, CPF4 and FS-3. These have the advantage that the hydrolysis of these unnatural substrates is easier to monitor directly. However, caution should be exercised when assessing the ATX inhibition with these compounds. Strikingly, we observed that the inhibitor S32826, which was reported as a nanomolar ATX inhibitor with an artificial substrate (Ferry et al. 2008), did not fully inhibit ATX when LPC was the substrate but inhibited it fully in the presence of an artificial substrate pNP-TMP (Pantsar et al. 2017a). In the original study, the S32826 was shown to inhibit radiolabeled LPC hydrolysis and we have no explanation for the observed discrepancy between these experiments. Furthermore, Joncour et al. reported both inhibition of ATX with FS-3 and LPC as a substrate (Joncour et al. 2017), whereas a poor correlation with these substrates is evident from our data. The tested compounds displayed clearly better IC\textsubscript{50} values with FS-3 as a substrate. Therefore, it is important to clearly identify the substrate when stating or comparing these activity values. Formerly, most of the reported inhibitors were reported with their IC\textsubscript{50} values with these non-natural substrates, which makes comparison of the inhibitors challenging. These discrepancies exemplify the risks of using unnatural substrates in the assay, and furthermore, obtaining false positive or false negative hits.

Finally, it is still unclear, in fact it has not even been investigated, if the inhibition of ATX's enzymatic activity is enough to result in therapeutic efficacy. This raises the question, does the binding of ATX to the activated β3-integrins itself cause a cellular response? This cellular response, if it exists, may be either positive or negative with regard to efficacy. A positive, enhanced effect, would be observed if there is a functional negative feedback-loop in the cellular signaling system, which would break the cellular signaling upon constitutive ATX binding to the integrins (if this negative feedback-loop is still intact and not already suppressed by the genetic alterations in the cancer). The negative effects would appear if simply ATX binding to the integrins would result in a positive cellular response. Moreover, as ATX is bound to the integrins, it obviously blocks the abilities of other proteins to bind at least to the occupied site. In addition, it has been shown that integrin β3 also displays ligand-independent signaling via the TGF-β pathway (Rapisarda et al. 2017). It is unclear how ATX binding to integrins would affect to the intracellular signaling complex associated with the integrins. The cytoplasmic domains of integrins form a basis for the assembly of large intracellular signaling complexes (Campbell and Humphries 2011). In fact, the catalytically inactive ATX or the plain SMB domains exhibit enhanced migration of cells, up to ~40% of migration effect of the ATX (Wu et al. 2014). This could imply that even if an inhibitor is able to achieve total inhibition of ATX enzymatic activity, it would still not hinder totally ATX’s ability to promote cell migration.

However, it may be possible to design an inhibitor which would, additionally to LPC hydrolysis inhibition, prevent the ATX binding to integrins. For instance, the inhibitor could change the ATX dynamics and/or conformation, thereby preventing the binding to integrins. However, this type of approach has its own risks. Excessive conformational changes in the protein induced by small-molecule inhibitors, could result in immune mediated adverse effects, especially as ATX is located in the circulation (Billheimer et al. 2002).

Moreover, there seems to be sufficient data that there are some specific cellular settings, for instance a specific mutation profile, where the ATX inhibition would be extremely beneficial. Furthermore, confirmation of this data would also clarify at which state of the disease one should best target ATX. Should the drug be delivered at all stages disease phase/state, or should it be introduced only in the latest stage, to prevent the metastatic potential of the tumor? Moreover, is it sufficient to inhibit ATX as a monotherapy, or does one need to combine some other agent with the ATX inhibitor?

Furthermore, as no ATX inhibitors have yet been clinically tested for cancer treatment, it is currently unknown what the optimal dose levels will be for cancer treatment, compared to those used in idiopathic pulmonary fibrosis. In other words, if a much higher dosage is needed, this could lead to toxicity related issues. In addition, it is currently unknown how long treatment periods with ATX inhibition would be needed. It may be that if long-term chronic therapy would
be required, the toxicity could become a major issue. For instance, wound healing could be one of the chronic toxicity related issues with ATX inhibition (Watterson et al. 2007, Lee et al. 2013).

In conclusion, as an extracellular secreted enzyme, ATX represents a highly attractive drug target. Still, as stated above, many issues about efficacy need to be resolved before one can discover a useful small-molecule agent; however, should this be achieved, the compound could be utilized in HCC treatment.
4.3 K-RAS

4.3.1 K-Ras as drug target
The gene \(KRAS\), located in chromosome 12, encodes the small GTPase K-Ras (K-Ras) 21 kDa protein. K-Ras is a crucial signal-transducing protein, which conveys signals from multiple sources and amplifies the signal to various target proteins, the effector proteins. It belongs to the superfamily of Ras proteins, which consists of 156 members. This family is further divided into subfamilies Arf (27), Rab (61), Ran (1), Rho (20), other (9) and Ras (36) (Wennerberg et al. 2005). The members of Ras-family most similar to K-Ras are the H-Ras and N-Ras proteins (Fig. 18). Moreover, there exist two splice-variants of K-Ras, K-Ras 4A and K-Ras 4B. These isoforms differ only in the C-terminal hypervariable region (HVR), where 4B displays a polybasic region and one site for the farnesylation and 4A, in addition to the farnesylation site, has one additional cysteine, which can be palmitoylated.

![Figure 18. Alignment of KRAS 4A, KRAS 4B, NRAS and HRAS. The proteins identical in their P-loop, switch-I and switch-II regions, with overall high similarity, except for their Hypervariable region (HVR). Alignment conducted with Clustal Omega 1.2.4 (Sievers et al. 2011).](image)

K-Ras activity is regulated through nucleotide binding and the intrinsic enzymatic activity; it binds GDP and GTP in its inactive and active states, respectively (Vetter and Wittinghofer 2001). K-Ras has intrinsic GTPase activity, i.e. it can hydrolyse the bond between the \(\gamma\)- and \(\beta\)-phosphate, resulting in GDP and K-Ras inactivation. However, this intrinsic ability is rather weak, and in this respect, the inactivation of K-Ras is mainly controlled by GTPase activating proteins (GAPs) that enhance this hydrolysis process by introducing an arginine finger to the site, thereby catalyzing the hydrolysis (Scheffzek et al. 1997, Hunter et al. 2015).

A GAP catalyzes the reaction by introducing an arginine finger into the site, which aids the hydrolysis reaction by stabilizing the negative charge in the leaving phosphate group. Surprisingly, there is scarce evidence of the K-Ras interaction with any of the listed GEFs and GAPs. One reason for this may be that most of the early examined Ras were mainly focused on H-Ras. It is worth noting that in the older literature, Ras proteins were usually described broadly as Ras proteins, and not identified as separate proteins, H-Ras, K-Ras or N-Ras. In fact, recently it was shown that they display discrepancies even in their biochemical properties (Johnson et al. 2017). Nevertheless, it is generally assumed that at least most of the Ras-family GAPs, bind to all Ras proteins (Bos et al. 2007, Vigil et al. 2010). These RasGAPs include: Ras GTPase-activating
protein 1 (RASA1, also known as p120GAP) (Hunter et al. 2015); Ras GTPase-activating protein 2 (RASA2); Ras GTPase-activating protein 3 (RASA3); Neurofibromin (NF1) (Cichowski & Jacks 2001); Ras GTPase-activating-like protein IQGAP1 (IQGAP1), which does not exhibit GAP activity (Maertens & Cichowski 2014); Plexin-B1 (PLXNB1); RasGAP-activating-like protein 1 (RASAL1); Ras GTPase-activating protein nGAP (RASAL2); RAS protein activator like-3 (RASAL3). However, clear evidence is missing if there is any specificity for K-Ras among the RasGAPs.

Correspondingly, the Ras is activated by an upstream stimulus. For instance, the activation of a growth factor receptor leads to an activation of a Ras guanine exchange factor (GEF) via GRB2 and/or SHC1-4 adaptor proteins. This GEF factor binds to Ras and displaces GDP from its binding site by opening the switch-I and pushing with its α-helix the ligand away from the site (Margarit et al. 2003). Many GEFs have been shown to bind K-Ras e.g. Son of sevenless homolog 1 (SOS1) (Chardin et al. 1993), Rap1 GTPase-GDP dissociation stimulator 1 (RAP1GDS1) (Mizuno et al. 1991), Ras guanyl-releasing protein 2 (RASGRP2) (Clyde-Smith et al. 2000) and Ras-GEF domain-containing family member 1A (RASGEF1A) (Ura et al. 2006). Moreover, evidence of H-Ras binding has been shown for RAS guanyl-releasing protein 1 H-Ras, RasGRP1 (Iwig et al. 2013), RAS guanyl-releasing protein 4 (RASGRP4) (H-Ras) (Reuther et al. 2002) and for Ras-specific guanine nucleotide-releasing factor 1 and 2 (RASGRF1, RASGRF2) H-Ras (Gotoh et al. 2001). However, data for the RasGEF family members such as Son of sevenless homolog 2 (SOS2), RAPGEF1-2 and RASGRP2-3 is missing.

K-Ras is a signal-transducer, which operates in the intracellular side of the plasma membrane. To conduct proper spatiotemporal signaling, K-Ras is targeted to the membrane by post-translational modifications, as it is unable to bind to membrane per se. These modifications, therefore, control K-Ras localization, and thereby its activity. The K-Ras is modified by lipid-anchors in its hypervariable region (HVR), which assist the binding to the membrane. Farnesyltransferase (FTase) farnesylates C186 and C185 in K-Ras4A and K-Ras4B, respectively. Interestingly, if the FTase is inhibited, K-Ras can overcome the inhibition by attaching a geranylgeranyl to the same position by Geranylgeranyltransferase type I (GGTase-I) (Rowell et al. 1997, Whyte et al. 1997). Moreover, K-Ras4A is palmitoylated at C180 by palmitoylacyltransferase (Buss & Shefton 1986, Aicart-Ramos et al. 2011). This is truly necessary to allow K-Ras4A to bind plasma membrane, as the monofarnesylated protein displays only a modest affinity for the membrane (Silvius and l’Heureux 1994, Shahinian and Silvius 1995). In contrast, K-Ras4B contains a polybasic region in its HVR, which helps it to bind to the negatively charged membrane. However, an orphan G-protein coupled receptor, GPR31, is needed to deliver K-Ras4B to the membrane after the farnesylation/geranylgeranylation (Fehrenbacher et al. 2017). The prenylation, farnesylation or geranylgeranylation, and the palmitoylation are conducted in intracellular compartments and the Golgi, respectively (Rocks et al. 2010). Moreover, K-Ras localization to the plasma membrane is regulated by phosphodiesterase δ (PDEδ) (Chandra et al. 2011). However, whereas K-Ras4B binds to PDEδ, K-Ras4A does not (Dharmaiah et al. 2016). In addition, K-Ras4B can be phosphorylated at S181 by protein kinase C (PKC), which neutralizes the positive charge in the polybasic region, and thereby attenuates K-Ras from the membrane (Bivona et al. 2006).

Furthermore, K-Ras can be ubiquitinated at K104 and K147 (Sasaki et al. 2011). The ubiquitination at K147 results in increased affinity to Raf and PI3K. Moreover, K-Ras4B activity can be controlled by acetylation at K104 (Yang et al. 2012). This acetylation was shown to interfere at least with the GEF-mediated GDP/GTP exchange, thereby reducing K-Ras activity. Furthermore, HDAC6 and SIRT2 are able to de-acetylate K-Ras, and thus reactivate it after this acetylation (Yang et al. 2013).

K-Ras expression may also be controlled through miRNAs. These miRNAs include miR-30c (Tanic et al. 2012), Let-7a (Wang et al. 2013) and miR-96 (Saud et al. 2014).

The K-Ras binds it effector proteins, mainly through its switch-I and switch-II regions (Fig. 19). When no effectors are bound to Ras, these switch-regions appear to be highly-mobile. We
observed this in our MD-simulations (Pantsar et al. 2018), and this is also observed in the NMR-studies of Ras (Spoerner et al. 2001, Stumber et al. 2002, Spoerner et al. 2004, Spoerner et al. 2010). Moreover, the switches are disordered in the crystal structures of K-Ras, where no crystal contacts appear in the regions (PDB IDs: 3GFT, 4DSN, 4EPR, 4L8G, 4LDJ, 4LPK, 4LRW, 4OBE, 4QL3, 4TQ9, 4TQA, 4WA7, 5UQW, 5TB5, 5US4, 5USJ).

Figure 19. Structure of the GDP-bound K-Ras. Highlighted in the structure is the Gly12 (orange mesh-surface) and its Ca-carbon location (orange, sphere). The effector protein binding interface forming switch-I (residues 30-40) and switch-II (residues 58-72) regions highlighted with red and blue, respectively. GDP displayed as stick representation and Mg2+-ion as sphere (magenta).

K-Ras has numerous effector proteins. These proteins use a ubiquitin (UB)-like fold to bind to K-Ras. With the best characterized Ras-effector proteins, the UB domain can be further classified into the Ras-binding domain (RBD) and the Ras-association domain (RA) (Nassar et al. 1995, Ponting & Benjamin 1996). The best characterized K-Ras effector proteins are listed here. To note, the list is not exhaustive with the three best characterized effector proteins being the Raf-kinases, phosphatidylinositol 4,5-bisphosphate 3-kinases (PI3Ks) and Ral guanine nucleotide dissociation stimulator (RALGDS, also known as RalGEF).

The Raf-kinases consist of three family members, all of which K-Ras is able to bind: RAF proto-oncogene serine/threonine-protein kinase (RAF1) (Rodriguez-Viciana et al. 2004, Hunter et al. 2015), Serine/threonine-protein kinase B-raf (BRAF) (Rodriguez-Viciana et al. 2004) and Serine/threonine-protein kinase A-Raf (ARAF) (Rodriguez-Viciana et al. 2004). Moreover, among the Raf-kinases, K-Ras affinity is clearly highest towards RAF1, and also BRAF and ARAF
affinities are diminished compared to N-Ras and H-Ras. With respect to the PI3Ks, K-Ras prefers the phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform (PIK3CA, PI3Kα, p110α). K-Ras has lower effects on signaling via p110γ and p110δ, and no effect at all on p110β (Rodriguez-Viciana et al. 2004). In addition to RALGDS (Vetter et al. 1999, Rodriguez-Viciana et al. 2004), K-Ras is also able to bind Ral guanine nucleotide dissociation stimulator-like 1 and 2, (RGL1, RGL2) (Rodriguez-Viciana et al. 2004). Furthermore, K-Ras has been identified to bind many proteins associated with distinct functions (Table 2). Moreover, K-Ras 4B binds to calmodulin, which inhibits the S181 phosphorylation (Villalonga et al. 2001, Alvarez-Moya et al. 2010).

K-Ras seems to share the ability to bind the same effector proteins as H-Ras (at least in vitro); for instance, AurkA (Umstead et al. 2017), PLCε (Kelley et al. 2001), RASSF5 (Vavvas et al. 1998), RGL (Kikuchi et al. 1994), RGL2 (Peterson et al. 1996) and RIN1 (Han & Colicelli 1995). This is not surprising, as the switch-regions in the binding interface are completely identical. Therefore, most likely, K-Ras is able to bind several effector proteins as has been encountered with for H-Ras. K-Ras binding to these proteins has not been confirmed to date; such as T-lymphoma invasion and metastasis-inducing protein 1 (TIAM1) (Lambert et al. 2002). It is noteworthy that H-Ras and K-Ras display different affinities to distinct effector proteins (Nakhaeizadeh et al. 2016). For instance, K-Ras shows higher affinity towards RalGDS-RA and lower affinity to RASSF5-RA compared to H-Ras. Moreover, the difference in membrane binding of K-Ras and H-Ras, and thereby their location on the membrane, exerts a major effect on the signaling specificity in vivo.

Importantly, it was reported that also the plasma-membrane dimer formation and the nanoclustering on the membrane affected K-Ras mediated signaling (Nan et al. 2015, Zhou et al. 2017).

These effector proteins link K-Ras to mitogen-activated protein kinase (MAPK), Hippo-, PI3K, and Ral-signaling pathways. By cooperating with these various effector proteins, K-Ras can control the cellular proliferation, survival, invasion, angiogenesis and metastasis. K-Ras also affects cellular energetics via metabolism and also has an effect on inflammation pathways. In other words, K-Ras is able to control the majority of the hallmarks of cancer (Hanahan and Weinberg 2011). Consequently, KRAS is now recognized as a classical oncogene along with MYC. KRAS is frequently mutated, especially in the pancreatic, colorectal and lung cancers (COSMIC v. 80, Forbes et al. 2017). Usually, the mutation is observed in position 12 (86%), although there are a few other hotspots causing oncogenic mutations i.e. in positions 13 and 61. Mutations of KRAS are especially common in PDAC.

Amplification of K-Ras is observed in not-so-traditionally K-Ras linked tissues: the testis 32%, and in the stomach 8% (COSMIC v.81, Forbes et al. 2017). In addition, the ovaries, which also is a frequent tissue with a K-Ras position 12 mutation, exhibit K-Ras amplification in 6% of the samples in the database.

The quantity of K-Ras proteins in cells varies extensively. In a study which quantified Ras proteins in an isogenic colorectal cancer cell line, SW48, with WT and mutated Ras, the wild-type cells contained approximately 150,000 K-Ras4B proteins per cell (Mageean et al. 2015). Additionally, the ratio of isoforms has been studied. It was reported that the ratio between the transcripts of the isoforms in various cancer cell lines was in favour of K-Ras4B over K-Ras4A (Tsai et al. 2015). This was especially observed with pancreatic cancer cell lines (PANC-1, G12D; BxPC3, WT; MiaPaCa-2, G12C), where the K-Ras 4B was clearly dominant. On the contrary, in colorectal tumor samples, the ratio was found to be more equal between the isoforms.
Table 2. K-Ras interacting proteins.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Full name</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFDN; AF6</td>
<td>Afadin</td>
<td>Linnemann et al. 1999, Rodriguez-Viciana et al. 2004</td>
</tr>
<tr>
<td>APBB1IP</td>
<td>Amyloid beta A4 precursor protein-binding family B member 1-interacting protein</td>
<td>Rodriguez-Viciana et al. 2004</td>
</tr>
<tr>
<td>ARAF</td>
<td>Serine/threonine-protein kinase A-Raf</td>
<td>Rodriguez-Viciana et al. 2004</td>
</tr>
<tr>
<td>AurkA</td>
<td>Aurora kinase A</td>
<td>Umstead et al. 2017</td>
</tr>
<tr>
<td>BRAF</td>
<td>Serine/threonine-protein kinase B-raf</td>
<td>Rodriguez-Viciana et al. 2004</td>
</tr>
<tr>
<td>Calmodulin</td>
<td>Calmodulin-1</td>
<td>Villalonga et al. 2001, Alvarez-Moya et al. 2010</td>
</tr>
<tr>
<td>GRB7</td>
<td>Growth factor receptor-bound protein 7</td>
<td>Rodriguez-Viciana et al. 2004</td>
</tr>
<tr>
<td>PI3Kα; p110α</td>
<td>phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform</td>
<td>Rodriguez-Viciana et al. 2004</td>
</tr>
<tr>
<td>PI3Kγ; p110γ</td>
<td>phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit gamma isoform</td>
<td>Rodriguez-Viciana et al. 2004</td>
</tr>
<tr>
<td>PI3Kδ; p110δ</td>
<td>phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit delta isoform</td>
<td>Rodriguez-Viciana et al. 2004</td>
</tr>
<tr>
<td>RAF1; CRAF</td>
<td>RAF proto-oncogene serine/threonine-protein kinase</td>
<td>Rodriguez-Viciana et al. 2004, Hunter et al. 2015</td>
</tr>
<tr>
<td>RALGDS; RalGEF</td>
<td>Ral guanine nucleotide dissociation stimulator</td>
<td>Vetter et al. 1999, Rodriguez-Viciana et al. 2004</td>
</tr>
<tr>
<td>RAPH1; ALS2CR9</td>
<td>Ras-associated and pleckstrin homology domains-containing protein 1</td>
<td>Rodriguez-Viciana et al. 2004</td>
</tr>
<tr>
<td>RASSF1A</td>
<td>Ras association domain-containing protein 1</td>
<td>Donninger et al. 2007, weak binding in Rodriguez-Viciana et al. 2004</td>
</tr>
<tr>
<td>RASSF2</td>
<td>Ras association domain-containing protein 2</td>
<td>Vos et al. 2003, weak binding in Rodriguez-Viciana et al. 2004</td>
</tr>
<tr>
<td>RASSF4</td>
<td>Ras association domain-containing protein 4</td>
<td>Eckfeld et al. 2004, Rodriguez-Viciana et al. 2004</td>
</tr>
<tr>
<td>RASSF5; NORE1</td>
<td>Ras association domain-containing protein 5</td>
<td>Rodriguez-Viciana et al. 2004</td>
</tr>
<tr>
<td>RASSF6</td>
<td>Ras association domain-containing protein 6</td>
<td>Allen et al. 2007</td>
</tr>
<tr>
<td>RASSF9; P-CIP1</td>
<td>Ras association domain-containing protein 9</td>
<td>Rodriguez-Viciana et al. 2004</td>
</tr>
<tr>
<td>RGL1</td>
<td>Ral guanine nucleotide dissociation stimulator-like 1</td>
<td>Rodriguez-Viciana et al. 2004</td>
</tr>
<tr>
<td>RGL2</td>
<td>Ral guanine nucleotide dissociation stimulator-like 2</td>
<td>Rodriguez-Viciana et al. 2004</td>
</tr>
<tr>
<td>RIN1; RIN2; RIN3</td>
<td>Ras and Rab interactor 1–3</td>
<td>Rodriguez-Viciana et al. 2004</td>
</tr>
<tr>
<td>PLCε</td>
<td>Phospholipase Cε</td>
<td>Rodriguez-Viciana et al. 2004</td>
</tr>
<tr>
<td>RAPGEF2; PDZGEF1</td>
<td>Rap guanine nucleotide exchange factor 2</td>
<td>Rodriguez-Viciana et al. 2004</td>
</tr>
</tbody>
</table>
4.3.2 Targeting K-Ras

Multiple strategies to target oncogenic Ras have been attempted to date with very varying levels of success. The traditional way to target the protein was at the ligand binding site and to mimic the inactive GDP, as this type of approach has been highly successful with other proteins such as the kinases. However, this approach was doomed in its infancy, as the natural ligands, GDP and GTP, display binding affinities towards Ras proteins in the picomolar range whereas the cellular concentrations of these molecules are in the millimolar scale (John et al. 1990). Due to these abundant high-affinity natural ligands the competition of the same binding site with a small-molecule is an unfeasible approach. Subsequently, the discovery of the FTase inhibitors was hoped to overcome these problems (Gibbs et al. 1994). However, the discovery of these inhibitors, such as lonafarnib and tipifarnib, had been mainly investigated in H-Ras cell and mouse models (Kohl et al. 1993, Kohl et al. 1995). It should be remembered that at that time, Ras proteins were considered as more or less identical. Moreover, mainly breast cancer models were used, where Ras mutations are normally rather rare (Rowinsky 2006). Prior to embarking on the expensive clinical trials, it was also shown that inhibitors did not have any effect on K-Ras or N-Ras membrane localization (Whyte et al. 1997), since K-Ras is able to overcome the FTase inhibition via GGTase-I (see section 4.3.1 for details of post-translational modifications). Therefore, in hindsight, it is not surprising that the developed inhibitors showed a lack of efficacy in clinical trials against pancreatic cancer (Cohen et al. 2003, Van Cutsem et al. 2004, Macdonald et al. 2005). Moreover, the dual inhibition of FTase and GGTase-I evoked severe toxicity, and this approach was terminated (Lobell et al. 2001).

After these initial disappointments, the Ras proteins were long considered as undruggable. This was due to the fact that there were no evident binding pockets available in the published Ras structures (apart from the GTP binding pocket). Next, as the knowledge of the Ras biology extended, the focus was shifted to targeting the downstream signaling pathways (Baines et al. 2011). Whereas the inhibition of a single signaling pathway downstream of Ras was not sufficient, dual-inhibition showed more promising results (Little et al. 2011, Alagesan et al. 2015). Dual inhibition is currently being investigated in clinical trials (Jokinen & Koivunen 2015). In addition to the effector pathway targeting, synthetic lethality screens for Ras have been conducted in order to find novel drug targets that could be effective in Ras-mutant harboring tumors (Luo et al. 2009).

Recently, some success in direct targeting of Ras proteins has been achieved. First, targeting the GEF-mediated nucleotide exchange was introduced. However, the inhibition of GEF-mediated nucleotide exchange has mainly focused on H-Ras and not on K-Ras (Maurer et al. 2012, Sun et al. 2012, Shima et al. 2013, Burns et al. 2014).

Perhaps, one of the most promising recent advances was the discovery of G12C targeting covalent inhibitors (Ostrem et al. 2013, Lito et al. 2016). This type of inhibitors, such as ARS853, attach covalently to the mutant cysteine, with rest of the molecule binding to a newly revealed pocket under the switch-II region (PDB ID: 5F2E) (Patricelli et al. 2016). Moreover, these compounds are only able to bind the GDP-bound K-Ras, and not the GTP-bound version. An alternative type of designed covalent inhibitors, such as SML-8-73-1, contains a GDP-scaffold and in this way, they can compete with GDP/GTP, as it occupies their binding site (PDB ID: 4NMM) (Hunter et al. 2014). SML-8-73-1, as it contains a biphosphate, is not capable of penetrating through the cell membrane (Hunter et al. 2017). Moreover, it was shown to bind, albeit to a lower degree, also within the small GTPases to the ARL3 from the Arf-family, and EFTUD1 and GUF1 from the Elongation factor-family. Most recently, covalent quinazoline inhibitors were described (Zeng et al. 2017). Although the covalent inhibitors are promising, the frequency of G12C mutants out of all K-Ras position 12 mutations is only 14% (Pantsar et al. 2018). Therefore, these inhibitors might be only truly useful in lung cancer, where this mutation is the dominant type, thereby excluding their rational use in PDAC.

Furthermore, KRAS-PDE8 interaction inhibitors have been developed (Zimmermann et al. 2013, Zimmermann et al. 2014, Papke et al. 2016). These inhibitors disrupt the localization of K-Ras to the cell membrane. The inhibitors have been shown to be effective in PDAC cell-lines.
In recent years, the direct inhibition of Ras and effector protein(s) interactions have been approached with several distinct strategies.

An effector protein blocking protein, a monobody, NS-1 was discovered (Spencer-Smith et al. 2016). It was shown to bind to the α4-β6-α5 region of Ras, resulting in disruption of Ras dimerization and nanoclustering, thereby inhibiting Raf-mediated signaling. Interestingly, NS1 binds to both H-Ras and K-Ras, but not to N-Ras.

A small-molecule Ras-mimetic that binds to effector proteins and prevents their binding to Ras was recently introduced (Athuluri-Divakar et al. 2016). Rigosertib, originally described as a polo-like kinase 1 (PLK1) inhibitor (Gumireddy et al. 2005), was shown to bind to RBDs of Raf-kinases, RalGDS and PI3Ks (Athuluri-Divakar et al. 2016). Interestingly, PLK1 was identified in the screen of the Ras synthetic lethal interactions (Luo et al. 2009). However, rigosertib does not inhibit PLK1 kinase activity (Steegmaier et al. 2007). It was shown to inhibit c-Raf phosphorylation at S338, which prevents its association with PLK1 and also c-Raf kinase activity (Athuluri-Divakar et al. 2016). Later, the effect of rigosertib on Ras signaling was suggested to be linked to the induction of mitotic stress via JNK, and not in direct inhibition (Ritt et al. 2016). Moreover, in view of this confusing biological data with rigosertib, it may not be surprising that also the possibility that rigosertib may be a non-specific protein alkylator was recently suggested (Archambault & Normandin 2017).

Moreover, there are antibody mimetics that specifically target GTP or GDP bound Ras on top of the switches (Guillard et al. 2017). The structures consist of Designed Ankyrin Repeat Protein (DARPin), and crystal structures of the complexes have been published (PDB IDs: 5O2S, 5O2T). However, the residence time for the GTP specific antibody was shown to be quite poor.

Furthermore, antisense-tactics against K-Ras were recently introduced. An antisense oligonucleotide, AZD4785, targets KRAS mRNA irrespective of the mutation codon sites (Ross et al. 2017). This confers the advantage that it is able to target all mutants; unfortunately, it also targets WT K-Ras as well. The effectiveness was demonstrated in immunocompromised xenograft and patient-derived xenograft mouse models. Moreover, the knockdown of KRAS by AZD4785 (74–94% depending on the tissue) seems to be well tolerated and did not result in any major adverse effects, at least in mouse and monkey models.

Moreover, a specific G12D mutant targeting RNA-interference (RNAi) therapy approach was recently reported (Kamerkar et al. 2017). Exosomes engineered to carry shRNA or siRNA have been developed; these have displayed effectiveness in multiple mouse models of pancreatic cancer.

Synthetic lethality targets have been also actively sought. A potential target, E3 ubiquitin-protein ligase SIAH2 (SIAH2), has been proposed to act as a “gatekeeper” in the K-Ras signaling pathway, and is currently being actively investigated (Van Sciver et al. 2017). Blocking the action of SIAH2 prevented Ras mediated transformation and tumorigenesis (Schmidt et al. 2007). SIAH2 is an E3 ubiquitin ligase, and in K-Ras tumorigenesis it may be crucial for the stability of specific signaling molecules and in aiding the cells to overcome K-Ras induced cellular stress (Van Sciver et al. 2017). Moreover, it was found to be a tumor specific biomarker in pancreatic cancer (Qin et al. 2015). However, no potent and selective molecule against SIAH2 has been reported to date. For instance, vitamin K3 (menadione), which binds covalently to SIAH2 and inhibits the ligase activity with an IC₅₀ value of 20 μM (Shah et al. 2009), binds also to SIAH1 (Zhang et al. 2017).

To conclude, significant progress in the field in K-Ras drug discovery has been accomplished to date. At present, the major trend seems to be on targeting directly the K-Ras-effector protein.
interactions. This seems a reasonable tactic, although there are conflicting results with rigosertib. Nevertheless, it is clear that the final words in the saga of the role of K-Ras in the treatment of cancers have still to be written.

4.3.3 The complexity of K-Ras position 12 mutations

Traditionally, the oncogenic properties of the Ras proteins have been linked to the deficiency of the GTP-hydrolysis (Gibbs et al. 1984) and the insensitivity for Ras-GAPs (Trahey and McCormick 1987). The mutant Ras proteins have been stated as being *constitutively active*. However, it has been disclosed that even mutant K-Ras undergoes GTP-GDP cycling (Lito et al. 2016). In fact, the mutant K-Ras proteins still possess some intrinsic GTP-hydrolysis activity (Hunter et al. 2015). It is known that even the mutants of the same position differ between themselves, and they display discrepancies in their biochemical characteristics. The intrinsic GTPase activity varies among the position 12 mutants; the G12C mutant still has ~70% GTPase activity compared to wild-type, whereas the G12D has ~30%, and the G12A, G12R and G12V have less than 10% of the wild-type activity. Nevertheless, all position 12 mutants are insensitive to the GTPase activity stimulating proteins (GAPs).

Moreover, the amount of mutant K-Ras with different mutations seems to vary in heterozygote cells (Mageean et al. 2015); where G12A, G12D and G12R exhibit mostly only mutant proteins, the G12V and G12S display almost half WT mutant proteins. In addition, K-Ras (wild-type or mutant) is present in greater quantities with G12D mutant, whereas the total amount of K-Ras with G12V mutation is similar to that occurring in the wild-type.

Furthermore, differences among position 12 mutations among the K-Ras mutants in cell cultures were first noted by Ihle et al. in 2012. They studied non-small cell lung cancer (NSCLC) cell-lines, and observed that the K-Ras G12D mutant activated PI3K and MEK signaling whereas G12C and G12V preferred the Ral-signaling pathway and displayed diminished EGFR mediated Akt activation. Correspondingly, the position 12 mutants differ with respect to their preference to bind specific effector proteins; for instance, the Raf-kinase affinity also varies among the mutants (Fig. 20). All of the position 12 mutants display diminished affinity towards CRAF (Hunter et al. 2015). The G12C displays ~85% affinity to Raf compared to the wild-type, G12A exhibits 45%, with the lowest affinity being displayed by G12D, G12R and G12V (~15–20%). Similarly, among PDAC cell-lines, we detected differences in the various K-Ras position 12 mutants among with respect to their signaling pathway specificities and dependencies with (Pantsar et al. 2018). Interestingly, the G12V mutant PDAC cell-lines displayed a dependency on the PI3K signaling pathway, and they were not affected with the Raf-signaling pathway inhibition.
In our MD simulations, we observed that the mutants display individual dynamics (Pantsar et al. 2018). Intriguingly, the mutants do not just differ with wild-type K-Ras in terms of their dynamics but also the mutants show discrepancies between themselves. This was especially clear with the MSMs generated for the G12D, G12R and G12V mutants. The different dynamics of mutants most likely result in discrepancy in their observed effector protein preference and downstream signaling.

Furthermore, we also discovered that the position 12 mutations show a tissue specific distribution that cannot be explained by mutation probabilities. Therefore, we speculate that these discrepancies observed in these mutants may explain their tissue specificity, i.e. a specific mutation may be more advantageous in a particular tissue or cellular environment.

Surprisingly, despite all these differences among the position 12 mutants, these discrepancies are not usually taken into account when conducting studies with mutant K-Ras.

Figure 20. The position 12 mutation affects to K-Ras binding affinity to CRAF-RBD. The wild-type K-Ras displays the highest affinity to CRAF (a). The data is also plotted as a comparison to the wild-type’s affinity (b) (The affinity data obtained from Hunter et al. 2015).
4.3.4 The efficacy problem with K-Ras inhibition

Efficacy is a major issue with K-Ras, as only a handful of candidate agents have been able to display any effect in in vitro or in vivo models to date. Although, K-Ras has been extensively studied during the last decades. The K-Ras oncoprotein have been stated as being constitutively active. Interestingly, some K-Ras mutations cause Noonan syndrome (e.g. V14I, T58I, D153V) (Schubbert et al. 2006). Similarly, as with the cancer related mutations, at least the V14I and T58I mutations exhibit defective intrinsic hydrolysis of GTP and are insensitive to GAPs; thereby leading to hyperactivity of K-Ras. Intriguingly, as this hyperactivity of K-Ras has been traditionally considered as the cause for K-Ras oncogenicity, these mutations that cause Noonan syndrome are not observed in cancer. This fact truly emphasizes the oversimplification of the old hypothesis that stated that the impaired GTP-hydrolysis would be the only reason to account for K-Ras oncogenicity.

This section will also focus on the potential efficacy caveats, especially related to the currently most promising developed inhibition strategies (see section 4.3.2). Moreover, the efficacy issues that are related to the discrepancies among the position 12 mutations (see section 4.3.3) will also be discussed.

With the exception of the covalent inhibition of K-Ras G12C and the antisense or siRNAs targeting K-Ras at the translational level, the recent approaches targeting K-Ras have mostly been targeting the effector protein–K-Ras interaction. Considering the plethora of the K-Ras effector protein interactions, these tactics are not likely to succeed if they cannot simultaneously hit all of the crucial effector proteins. Moreover, the affinities with these reported molecules targeting the effector protein – K-Ras interaction are quite low. With low affinity values and the plethora of the interacting proteins, it appears unlikely that one can achieve efficient signaling inhibition. Furthermore, the nonselective therapies that would target e.g. H-Ras and K-Ras regardless of the mutations, seem unlikely to be successful in a clinical setting, as these proteins are ubiquitously expressed and on their binding interface, they appear to be similar as wild-type protein in normal cells. Nonetheless, it is interesting to speculate, that overall K-Ras inhibition may be tolerable, since at least the siRNA treatment for a few weeks was not overtly toxic (Pecot et al. 2014, Ross et al. 2017).

Furthermore, especially with therapies targeting a specific effector protein, the development of resistance is a serious risk. As observed by us and others, at least some of the position 12 mutants are not dependent on individual downstream pathways (Alagesan et al. 2015, Junttila et al. 2015, Pantsar et al. 2018). Interestingly, even as the PI3K pathway seems to be the key pathway in PDAC development (Eser et al. 2013), sufficient efficacy is not observed when inhibiting this pathway alone.

The G12C mutants represent a rather high proportion of all known K-Ras position 12 mutations, 14% (COSMIC v. 80, Forbes et al. 2017, Pantsar et al. 2018). Nevertheless, in other tissues with the exception of the lung, but especially in the pancreas, the observed frequency of the G12C is rather low. Therefore, the main advantages with these inhibitors may be only achieved with lung cancers, where the frequency of this mutation is high as smoking is observed to promote this specific mutation (Dogan et al. 2012). The covalent inhibition against other K-Ras mutants appears to be unachievable, as the other mutants do not share suitable characteristics for covalent attachment as the cysteine. Furthermore, the other cysteines in K-Ras, such as C118, may not be targeted due to the similarities with the other Ras-proteins (Fig. 18). As the other Ras proteins contain the same cysteine, targeting this amino acid would lead to targeting of all Ras proteins. Overall, the efficacy problem with the covalent inhibitors is that they target only a subpopulation of K-Ras mutants. This raises an important question for the efficacy of the G12C covalent inhibition—how fast can a subpopulation of the tumor acquire another type of K-Ras mutation? As the therapeutic agent is specific (and dependent) solely on the presence of the G12C mutant, it will lose its efficacy should the tumor acquire a different K-Ras mutant subclonal population. Therefore, this gain-of-mutation would lead to an instant relapse.
Furthermore, with respect to efficacy, it is crucial to identify if a tumor is exhibiting multiple K-Ras mutations or not. Fortunately, at least in PDAC, the intratumoral heterogeneity of K-Ras seems to be rather rare (Hashimoto et al. 2016). This hints to the fact that the K-Ras mutation is a clonal mutation, which is a hint that it was an early incident in the evolution of the pancreatic tumor. However, a crucial question is will there exist some subclonal tumor cell population, which would be able to obtain a different K-Ras mutation? Some evidence for this kind of heterogeneity was observed emerging from patient derived pancreatic cancer xenografts (Pedersen et al. 2017). Thereby, the mutation-specific therapies may be at a risk for a relapse during the therapy. Although, the tumor evolution can be altered in patient derived xenograft models (Ben-David et al. 2017).

It is not simply with respect to the covalent inhibition of G12C, but also as the mutations in position 12 are not equivalent (see section 4.3.3), it is crucial to identify which mutation is harbored in the tumor. Moreover, the selection for the right mutation-harboring patients in clinical trials is crucial with mutant specific therapies. Therefore, during the diagnosis, the mutation status should be evaluated. Nowadays, however, usually only a single sample and/or a single-timepoint assessment is conducted when analyzing mutations, and the tumor evolution is neglected with this approach (Turajlic and Swanton 2017). Therefore, the tumor evolution should be monitored. Novel methods such as liquid biopsies by cell-free plasma DNA (cfDNA) can be utilized to monitor the tumor burden (Meador & Lovly 2015). Moreover, the circulating tumor cells CTCs (Krebs et al. 2014), can be used to monitor treatment resistance (Esposito et al. 2016). However, with CTCs, the K-Ras mutation status may not match the primary tumor’s status very well (Kulemann et al. 2017). Therefore, caution should be applied when using these methods. Consequently, the detection of the mutant in the clinical setting would be crucial when targeting K-Ras with therapies that are dependent on the specific mutation. First, the mutation type needs to be confirmed in the diagnosis, and it needs to be further reliably monitored in a clinical setting. Finally, the tests need to be capable of monitoring the K-Ras mutation heterogeneity reliably.

As the position 12 mutants display discrepancies in multiple properties, especially in the preference and dependence of their cellular signaling pathway, it is crucial to note that in order to achieve sufficient efficacy, it may be that different mutants need to be targeted differently. In other words, if a specific therapy displays efficacy against the G12D mutation, it may be that the same therapy would not display this efficacy against the G12V mutation. Therefore, with respect to the efficacy, it is crucial to completely identify and be aware of the individual differences among the mutants.

To complicate the issue further, even the copy number of K-Ras mutant has an effect on cellular signaling. In other words, if the cell has still one allele of wild-type K-Ras left, this will exert a significant effect on the signaling outcome. It was shown that the K-Ras allelic imbalance has an effect on the cellular signaling, tumor fitness and therapy resistance (Burgess et al. 2017). Therefore, also the allelic imbalance status has implications in the drug therapy response. Furthermore, it has been shown that homozygote or heterozygote K-Ras mutants exhibit distinct metabolic profiles (Kerr et al. 2016). Therefore, even the K-Ras mutation status (homozygote or heterozygote) have an influence on the cellular signaling, and thereby may affect a therapy’s efficacy.

It is noteworthy that in addition to position 12 mutants, also the G13D mutant exists, with an overall frequency of almost 15% of all K-Ras mutations. The mutation in the position 13 differs tremendously from position 12 mutations, as reflected in its faster nucleotide exchange kinetics (Hunter et al. 2015). Therefore, it is not indeed clear whether a similar strategy as applied in targeting position 12 K-Ras mutants could also be utilized with position 13 mutants. Thereby, highlighting the importance of identifying the specific mutant prior to any therapy utilization, and the outcome analysis.

Another essential question is if the tissue of origin needs to be taken account when targeting K-Ras. It is not only the position 12 mutants that differ from each other in their cellular signaling,
but also the tissue and the cellular environment has its own influence on the signaling. We observed a discrepancy in the position 12 mutant tissue distribution (Pantsar et al. 2018). Therefore, it may need to be considered if the inhibition/targeting of a specific mutant (e.g., G12V), has to be utilized with different tactics, when targeting the same mutant in a different tissue.
5 Discussion

The current drug therapy possibilities for patients with solid tumors are clearly inadequate. In fact, it has been noted that there is a minor overlap between cancer drivers and current cancer drug targets (Santos et al. 2017) and this seems to account for the poor outcomes for the patients. Therefore, there is an evident need for novel, and most importantly efficient, therapeutic options. Moreover, most of these novel drug candidates have been unsuccessful in clinical trials due to their lack of efficacy. Therefore, the potential efficacy related problematics in the studied potential drug targets is emphasized in this thesis. These three different drug targets covered in this thesis, each represent a unique efficacy-problem profile and highlighting the fact that the efficacy problem needs to be handled individually for different targets. Consequently, one cannot envisage that there will be one single general golden rule that would solve all the efficacy related issues for all drug targets. On the other hand, some similarities related to the efficacy problem among these targets are identifiable.

Quite often orthogonal assays are employed to confirm the compound activity in vitro. However, although being a reasonable and useful approach, this actually only solidifies the result of the first assay and does not fill or narrow the gap with the molecular events observed on the larger, cellular scale. Therefore, parallel metrics for the in vitro evaluation of the compounds should be utilized more often. A good example of this approach is that we saw totally different responses on the cellular scale with different AurkA inhibitors, and that these could be linked to the AurkA–Myc interaction (Dauch et al. 2016). This observed efficacy difference could not have been predicted from the kinase inhibition metrics of the compounds. Obviously, this does not mean that the kinase inhibition metrics should be neglected, as the kinase inhibition for an effective AurkA inhibition is clearly acknowledged.

Therefore, it is most important to identify the most crucial events for the efficacy of the compound at the molecular level, and to monitor these early enough. This would most likely to lead to improved transition rates in the clinical trials, as the “rotten apples”, which display equivalent properties e.g. kinase inhibition at the same level, would be discarded early on. Presumably, these additional assays in the early drug discovery phase would lead to increased costs and also take more time. Nevertheless, it is likely that there would be savings from the unnecessary clinical trials, which is the most expensive part of the drug discovery process. Furthermore, most importantly, this could lead to a higher success rate in the clinical trials and thereafter to the beneficial therapeutic options for the patients.

Moreover, this exemplifies the fact that one needs to understand the complexity of the cellular signaling and identify the most crucial components related to the specific environment. In addition, this means that in the future one may need to abandon the “one target – one solution” philosophy. Instead, one type of a drug molecule acting specifically against a particular target, may be effective only in a particular cellular environment, whereas another drug against the same target could be efficient in another environment. For instance, the dependence for the AurkA–Myc interaction was only displayed in the TP53 altered HCC environment (Dauch et al. 2016); therefore, the conformational change inducing inhibitors that distort this specific interaction, would only be effective in this specific type of HCC. Moreover, it is not so clear if this induced conformational change would be beneficial in non-Myc dependent cellular environment. In this non-Myc dependent environment, it may even be that these inhibitors would be significantly worse compared to the other inhibitors.

Additionally, one needs to pay attention to the plausible drug resistance mechanisms, which may emerge during the drug-therapy and how to detect and address to these risks (Holohan et al. 2013, Schmitt et al. 2016). Obviously, this is a crucial aspect in the drug’s efficacy. If the novel
drug therapy can be easily circumvented by the tumor, and there are no therapy options available against this resistant subclonal tumor population – little benefit will be acquired with the drug for the patient. Therefore, the drug’s target and the mechanism of action need to be evaluated thoroughly. The drug’s action should not be easily circumvented by the tumor; for instance, if the tumor could circumvent the drug’s effect by a selected mutation or a gene amplification that could be predicted to occur relatively fast. Alternatively, the drug should guide the subclonal evolution mechanism of the tumor into a dead-end that one can target with an available therapy. This seems to be somewhat overlooked, especially with some of the developed K-Ras targeted novel therapies. This is probably due to the reason that there has been a lack of any efficacy against mutant K-Ras harboring cell-lines or in vivo models.

As a younger drug discovery target, the caveats in efficacy with ATX are not so easily identified. This is probably due to the lack of available data. In my opinion, there is one crucial question in targeting ATX: is the enzyme activity inhibition in itself sufficient to achieve high-enough efficacy? An alarming example is observed with the inhibition of A urkA kinase. There is a lack of knowledge in the dynamical characteristics of the ATX, and it has been badly understudied. In fact, only one MD simulation study of ATX has been published (Koyama et al. 2012). In that study, the role of the glycosylation and nuclease domain for the catalytic activity was investigated in short 300 ns simulations. In the future, it would be important to investigate the dynamics of ATX and integrin interaction. Moreover, the potential effect of an inhibitor to the SMB domain dynamics should be scrutinized. As this could be an important aspect with respect to gaining reasonable efficacy, if the enzymatic activity inhibition is not solely enough for the ATX inhibition with true efficacy.

Although the three drug targets handled in this thesis appear first as unrelated and distinct proteins, they all are actually linked together in the complexity of the cellular signaling via the Ra lA protein. ATX mediates the cellular migration and invasion via the Ras pathway, resulting in an increase in the Ras activity (Aziziye h et al. 2009, Li et al. 2009). Moreover, the RasA is a critical factor for A urkA mediated motility and transformation (Wu et al. 2005). Interestingly, RasGAP, which shuts off the active RasA, was identified as a tumor suppressor in HCC (Kodama et al. 2016). Furthermore, an A urkA inhibitor combined with a miRNA, which targets RasA, displayed efficacy in a prostate cancer model (Epis et al. 2017). In addition, K-Ras seems to regulate A urkA overexpression by FOSL1 (Vallejo et al. 2017), adding another link between these proteins. The overexpression of A urkA was identified to potentiate H-Ras oncogenesis long before the discovery of the A urkA–H-Ras direct interaction (Tatsuka et al. 2005, Umstead et al. 2017). It remains to be seen if a direct link between K-Ras and A urkA can also be established, as K-Ras regulates A urkA expression (Dos Santos et al. 2016) and A urkA inhibition displays efficacy in the K-Ras G12D PDAC model (Xie et al. 2017).

Furthermore, to date and in the future, the drug targets will consist of increasing awareness of the importance of preventing protein–protein interactions (Scott et al. 2016), and not only traditional straightforward enzyme inhibition. Probably because of the oversimplification of the simple assays and their limitations, such as kinase inhibition assays which do not reflect to the complexity of the disease, the popularity of phenotypic based screens have undergone a resurgence (Moffat et al. 2017). The authors emphasize the complementary role of these phenotypic assays with the traditional target-based drug discovery. This makes sense, as seen from the three-distinct drug-target examples in this thesis, there are numerous interacting protein–protein interactions with the targets, and naturally one is unable to analyze all of these interactions individually. However, if one has some information of the most crucial interacting partner, this should be also monitored; for instance, the A urkA–MYC interaction with the A urkA inhibitors as well as the traditional kinase inhibition assay.

The putative model for the A urkA–Myc interaction that we generated helped to interpret the observed differences among different A urkA inhibitors. However, the model is not perfect and we are still uncertain how close to the real-life situation the model reflects. In fact, the putative model could have been further validated with MD simulations. However, due to time limitations,
this was not conducted. Future studies should validate and challenge our generated model. It would be advantageous to obtain experimental structural data of the complex. However, it may be that the crystallization of the complex is not possible due to the phosphorylation of the Myc that is needed for the binding and the complex may be dynamic. Therefore, one can predict that this experimental structural data confirmation may be achieved via NMR. Moreover, the exact knowledge of the details of this AurkA–Myc interaction may be crucial for the success of efficient AurkA-inhibitor development, and thereby their efficacy.

The interpretation of the crystal structural data of protein–protein interactions may be misleading. The interfaces may be formed due to the crystal contacts and do not represent the actual biological interfaces (Kobe et al. 2008). Moreover, the allosteric effects to the protein may not appear in crystal structures, as they do not necessary require a conformational change of a protein (Tsai et al. 2008, Tzeng & Kalodimos 2012, Rodgers et al. 2013). In fact, according to our MD simulations, this allosteric effect comes into play with position 12 mutations (Pantsar et al. 2018), where there are no major differences among their crystal structures. Moreover, this is highly-relevant also with ATX and AurkA. The allosteric effects of the inhibitors may be crucial with ATX inhibitors to the integrin–ATX interaction; with the AurkA inhibitors, this is definitely with the Myc–AurkA interaction.

Interestingly, the drug – target residence time concept has not been a point of focus with these three different drug targets, or at least, very little data has been reported publicly. The residence time may play a crucial role in drug’s efficacy (Lu et al. 2009, Sykes et al. 2009, Guo et al. 2012, Pan et al. 2013), and currently the sufficient residence time for each of these three targets is unknown. Therefore, more focus should be put on this aspect in future. Researchers have recently started to address the binding kinetics (the residence time prediction) with computational methods by applying molecular dynamics (Casasnovas et al. 2017, Tang et al. 2017b) Clearly, the residence time per se is not an enlightening factor for efficacy, if other important parameters are simultaneously neglected (Folmer 2017). Therefore, as useful as the residence time is, and it should be monitored, the molecules should not be evaluated exclusively by this parameter.

Furthermore, the tumors in the same organ among all the patients are not usually identical. Correspondingly, understanding the specific subtypes of the tumors of particular tissue, will aid identifying sensitivity against drugs (Senses et al. 2017) Moreover, the intratumoral heterogeneity and the tumor evolution (Rosenthal et al. 2017), should be noted during disease progression as well as during the therapy. Tumor heterogeneity in HCC among different patients and within the tumor nodules in a single patient (Craig et al. 2017). In addition, caution should be made when translating results from one tumor type to another, as tissue and cell-type specific differences in the oncogenic pathways and tumorigenesis exist (Schneider et al. 2017).

Researchers in Pfizer suggested three pillars that they claimed would define the probability of a drug candidate to succeed in the phase II to phase III transition (Morgan et al. 2012): These pillars are the exposure at the target site of action for a desired period of time; binding to the pharmacological target as expected for its mode of action; expression of pharmacological activity commensurate with the demonstrated target exposure and target binding. Moreover, in addition to these pillars, the thorough characterizations of the protein target and the candidate drug molecule are obligatory (Bunnage et al. 2015). Especially, one needs to understand the functional consequences to the target after the molecule binds to it. This is perhaps the most laborious task if one strives to achieve a proper understanding of all the consequences due to the vast amount of interactions that are displayed by a typical drug target. However, now we are approaching the point in drug discovery that will shed light on this issue, as the necessary technologies that can be utilized have made major advances towards this ultimate goal.

The gap between the molecular scale world and the cellular scale (even to the organ scale) has narrowed tremendously. This can be attributed to technological advances related to the early phase drug discovery, the emergence of various biophysical methods with the more exact biochemical and biological measurements, which are already playing a major role in the early phase of drug discovery (Renaud et al. 2016, Genick & Wright 2017). In other words, the real-
world phenomena are becoming more conducive to molecular modeling, as the rigorous experimental assays are able to produce high-quality data for the in silico models. A clear indication of this ‘technological enlightenment’ is that there is a growing trend that some cancer targets that were thought of as “undruggable” are now considered more as ‘difficult to drug’ or ‘yet to be drugged’ targets (Lazo & Sharlow 2016, Dang et al. 2017).

The molecular modeling has been successfully utilized in various drug discovery processes. It has been able to produce novel active molecules and enhance the binding affinity of the molecules. However, molecular modeling has not been able to answer the question, how does the molecule affect the target protein, i.e. how does a bound drug molecule affect to the target protein’s dynamics which further has an effect to the drug target’s protein–protein interactions, and thereby has a great influence on the drug molecule’s efficacy. In the future, molecular modeling should be able to answer this question, and further help to explain and predict the efficacy of the designed molecules. This would save resources as the bad candidates would be discarded earlier, reducing the unnecessary in vitro and in vivo testing, and most importantly, it could lead to reduced drug attrition rates in the clinical trials. However, because of the improved accuracy of force fields and the increased computational capacity, the MD simulations can now be conducted over meaningful time scales relating to the protein dynamics (see section 3.2.2), which were previously inaccessible. Nowadays, MD simulations are already an essential part of the structural elucidation of the drug-target, and have become crucial in the structure-based drug design and discovery (Bermudez et al. 2016). In the future, MD simulations will probably go beyond even this stage, and may be able to predict the effects induced by the drug molecule on the target protein. This would lead to much clearer efficacy prediction, when combined with the biophysical and the biochemical methods.

Furthermore, additional novel methods, which are extremely useful for the drug discovery, are constantly emerging. It is worth noting that the gap between the isolated protein and cellular assays is now bridged by the cellular thermal shift assay (CETSA) (Martinez Molina et al. 2013, Martinez Molina & Nordlund 2016). This assay is able to confirm the compound’s engagement with the target in living cells. Moreover, single-cell sequencing may aid in unraveling the complexity of cancer pathology (Baslan & Hicks 2017). In addition, the live-cell phenotypic biomarkers may help to target the optimal patient population and thus they can also aid in drug discovery (Sant et al. 2017). Moreover, the utilization of non-traditional, non-small molecule approaches is emerging for challenging drug targets (Valeur et al. 2017). It remains to be seen to what extent these will be successful at targets unsuitable for small-molecules.

To summarize, technological advances have improved the tools applied in the early phase of drug discovery to such an extent that we may be said to be entering into a new era in drug discovery where we will be able to identify and understand how a drug compound affects its target. This will eventually lead to a better understanding of the target and the disease, resulting in better efficacy in future clinical trials. Today, the researchers are no longer proceeding blindfolded along some route, but have arrived at the point where different disciplines are intersecting and shedding light on the road ahead.
6 Epilogue

This thesis was started with a quote from John McTiernan’s Predator (1987). That quote is referring to the ability to believe in success even in face of a formidable enemy, whose form, aims and technique are incomprehensible. The bleeding refers to vulnerability, in our analogy some identified weakness of a tumor. This can emerge as a potential drug target is discovered from a shRNA screen. Even after the discovery of a potential weakness in the tumor, from these three totally different potential drug-targets covered in this thesis, it is obvious that it is still a distant hope that it will be able to be targeted with a small-molecule that is really effective. First, at the onset, usually there are only ineffective weapons available in the arsenal. Therefore, one needs to create new weapons (drug molecules) from scratch. The plethora of interactions that the potential drug-targets exhibit does not make this task any easier, as these will determine what type of a weapon is truly effective against the target.

In order to achieve the ultimate goal of developing effective drugs against these currently untreatable solid tumors, the role of basic research cannot be sufficiently emphasized. Especially, in these times when, at least in Finland, there is an ever-increasing trend towards applied research, while simultaneously basic research has become more and more neglected. In order to diminish the gaps between the atomic–molecular, molecular–biological and biological–clinical levels, basic research is obligatory and cannot be bypassed. Without this basic research, the applied researcher will be approaching crucial intersections with blindfolds. The narrower that these gaps can become, the easier it will be to derive correct and precise conclusions from the obtained results, and further achieve success in the drug discovery program.

The human mind has a remarkable ability to create order by organizing and categorizing data. The more ordered and organized the data, the easier they will be to comprehend. However, the real world is not fixed, it is highly dynamic. This is especially true with a cancer, since tumor cells undergo rapid evolution and possess dynamic characteristics. Therefore, it is not surprising that it seems to be so hard to fight effectively against cancer, as the tumor appears to be, to some extent, beyond our comprehension. Therefore, I think the best way to conclude this thesis is a quote from the Ridley Scott’s Alien (1979), which is also, unfortunately, highly-applicable for the solid tumors.

“YOU STILL DON’T UNDERSTAND WHAT YOU’RE DEALING WITH, DO YOU? PERFECT ORGANISM. ITS STRUCTURAL PERFECTION IS MATCHED ONLY BY ITS HOSTILITY.”

IAN HOLM (ASH) – ALIEN (1979)
7 References


Archambault, Normandin K: Several inhibitors of the Plk1 Polo-Box Domain turn out to be non-specific protein alkylators. Cell Cycle 16: 1220-1224, 2017


Cichowski K, Jacks T: NF1 Tumor Suppressor Gene Function: Narrowing the GAP. Cell 104: 593-604, 2001


Harpole TJ, Delemotte L: Conformational landscapes of membrane proteins delineated by enhanced sampling molecular dynamics simulations. Biochim Biophys Acta 1860: 909-926, 2018


Jalink K, Moolenaar WH, van Duijn B: Lysophosphatidic acid is a chemoattractant for Dictyostelium discoideum amoebae. Proc Natl Acad Sci USA 90: 1857-1861, 1993


Li DW, Brüschweiler R: NMR-based protein potentials. Angew Chem Int Ed Eng 49: 6778-6780, 2010


Linnemann T, Geyer M, Jaitner BK, Block C, Kalbitzer HR, Wittinghofer A, Herrmann C: Thermodynamic and kinetic characterization of the interaction between the Ras binding domain of AF6 and members of the Ras subfamily. J Biol Chem 274: 13556-13562, 1999


Littlepage LE, Ruderman JV: Identification of a new APC/C recognition domain, the A box, which is required for the Cdh1-dependent destruction of the kinase Aurora-A during mitotic exit. Genes Dev 16: 2274-2285, 2002


Pantsar T, Rissanen S, Dauch D, Laitinen T, Vattulainen I, Poso A: Assessment of mutation probabilities of KRAS position 12 mutants and investigation of their long-time scale dynamics by atomistic molecular simulations and Markov state modeling. Submitted, 2018


Pearlman D, Case D, Caldwell J, Ross W, Cheatham T III, DeBolt S, Ferguson D, Seibel G, Kollman P: AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and


Rodriguez-Viciana P1, Sabatier C, McCormick F: Signaling specificity by Ras family GTPases is determined by the full spectrum of effectors they regulate. Mol Cell Biol 24: 4943-4954, 2004


Rowinsky EK: Lately, it occurs to me what a long, strange trip it’s been for the farnesyltransferase inhibitors. J Clin Oncol 24: 2981-2984, 2006


Sant GR, Knopf KB, Albala DM: Live-single-cell phenotypic cancer biomarkers-future role in precision oncology? NPJ Precis Oncol 1: 21, 2017


Shahinian S, Silvius JR: Doubly-lipid-modified protein sequence motifs exhibit long-lived anchorage to lipid bilayer membranes. Biochemistry 34: 3813-3122, 1995


Zhao ZS, Lim JP, Ng YW, Lim L, Manser E: The GIT-associated kinase PAK targets to the centrosome and regulates Aurora-A. Mol Cell 20: 237-249, 2005


TATU PANTSAR

The current drug therapy possibilities for patients with solid tumors are inadequate, and the drug candidates tend to fail in the clinical trials due to the lack of efficacy. In this thesis three promising drug targets, AurkA, ATX and K-Ras, were evaluated from the perspective of early phase drug design in the respect of their potential efficacy caveats. This thesis provides novel insights to the efficacy problems among these targets and how these problems could be addressed in the early phase drug design, which could lead to lower drug attrition rates related to the lack of efficacy.