**Thesis for the degree of doctor of philosophy in natural science, Specialization in chemistry**

# Computer-aided drug design approaches in developing anti-cancer inhibitors

## Chunxia Gao

**Department of Chemistry and Molecular Biology**
**Univerisity of Gothenburg**
**Gothenburg**
**2016**

Computer-aided drug design approaches in developing anti-cancer inhibitors

Chunxia Gao

Department of Chemistry and Molecular Biology University of Gothenburg
SE-412 96 Göteborg
Sweden

*Science is what we have learned about how to keep from fooling ourselves.*

*-Richard Feynman*

*To my beloved family*

# ABSTRACT

The thesis entitled "computer-aided drug design approaches in developing anti-cancer drug" is divided into a total of six chapters. In the first chapter, an overview of drug discovery and development are introduced. Nowadays, drug discovery and development has clearly changed the course of the disease treatment, numerous of therapeutic agents have been designed to treat cancers, cardiovascular diseases, infections etc. However, the whole process of drug discovery and development is lengthy and complicated, which require huge input of time, money, and resources. The whole process can be divided into five main stages, including target identification, lead discovery and optimization, preclinical tests and clinical trials. Acknowledging that drug discovery is complicated in general, oncology has one of the poorest records for developing into clinical trial phases, however, we have moved into a new "golden era" for cancer drug development through molecular targeting recently. In the second chapter, computer-aided drug design (CADD), which is effective methods for facilitating and expediting drug discovery and development process, are introduced. Some approved drugs that credited their discovery in large part to the tools of CADD were reported, and the application of CADD has extended to two directions in drug discovery and development: upstream for target identification and validation, and downstream for preclinical study, mostly ADME/T prediction. Though CADD holds a great promise for future progress in drug discovery and development, it is still an evolving technology and has number of limitations that should be resolved. In the third chapter, a detailed introduction of the anti-cancer targets in this thesis is given. The targets include tyrosine kinase RET, kinesin Eg5 and KIF18B, histone transferase Tip60 and the GTPase K-Ras. The involving signalling pathway, the cancer induced mechanism and the current available inhibitors for the targets are discussed. In the fourth chapter, the methodologies applied for different projects will be discussed, including homology modelling, docking, molecular dynamics simulations, structure-based pharmacophore, ligand-based pharmacophore and 3D-QSAR, and density functional theory. The detailed theoretical background and the main steps involve in the use of these methods are outlined. In the fifth chapter, I provide a summary of seven papers or manuscripts, which are related to this thesis. Based on the RET target, we have studied the interaction and pharmacophore of four DFG-out inhibitors, mutations in the cadherin like domain, and prediction and calculation of a photo-switchable inhibitor. An inhibitor was designed and validated for Tip60. Biphenyl-type inhibitors targeting Eg5 were fully investigated from

different perspectives. Inhibitors were designed to target the α4/α6 allosteric pocket of KIF18B, in order to investigate the specific molecular functional roles of KIF18. Mutations on the binding interface of K-Ras and GAP were looked into and the continuous active oncogenic signalling caused by specific mutations was explained. In the final chapter, conclusions and future perspective is given, CADD is indeed a very useful tool for pharmaceutical companies and academic research groups to search for potential drug candidates, meanwhile, it is also necessary to improve the current CADD methods. The different projects included in this thesis have the potential for further development.

# LIST OF PUBLICATIONS

**Paper I.**

*Characterization of interaction and pharmacophore development for DFG-out inhibitors to RET tyrosine kinase.* **Chunxia Gao,** Morten Grotli, and Leif A. Eriksson. *Journal of molecular modelling* (2015) 21:167 DOI: 10.1007/s00894-015-2708-z

**Paper II.**

*Defects in the calcium-binding region drastically affect the cadherin-like domains of RET tyrosine kinase.* **Chunxia Gao**, Morten Grotli and Leif A. Eriksson. *Physical Chemistry Chemical Physics (2016)* DOI: 10.1039/C6CP00042H

**Paper III.**

*Mechanism of action of a photo-switchable inhibitor targeting RET tyrosine kinase* **Chunxia Gao**, Morten Grotli and Leif A. Eriksson. Manuscript in preparation

**Paper IV.**

*Rational design and validation of a Tip60 histone acetyltransferase inhibitor.* **Chunxia Gao**, Emer Boruke, Martin Scobie, Melina A Famme, Tobias Koolmeister, Thomas Helleday, Leif A. Eriksson, Noel F. Lowndes and James A.L.Brown. *Scientific Reports* (2014) 4:5372 DOI: 10.1038/srep05372

**Paper V.**

*Analysis of Biphenyl Type Inhibitors Targeting the Eg5 α4/α6 Allosteric Pocket*. **Chunxia Gao**, Noel F. Lowndes and Leif A. Eriksson. Submitted

**Paper VI.**

*Designing inhibitors targeting KIF18B.* **Chunxia Gao**, Noel F. Lowndes and Leif A. Eriksson. Manuscript in preparation

**Paper VII.**

*Impact of mutations on K-Ras-p120GAP interaction.* **Chunxia Gao**, and Leif A.Eriksson. *Computational Molecular Bioscience* (2013) 2:3 DOI: 10.4236/cmb.2013.32002

# RELATED PUBLICATIONS NOT INCLUDED

**Paper VIII.**

*Exploration of multiple Sortase A protein conformations in virtual screening*. **Chunxia Gao**, Ivana Uzelac, Johan Gottfries and Leif A. Eriksson. *Scientific Reports* (2016) 6:20413 DOI: 10.1038/srep20413

**Paper IX.**

*Morin inhibits Sortase A by allosteric binding in the dimerization interface.* Ivana Uzelac, **Chunxia Gao**, Thomas Jacso, Patric M. Wehrli, Thomas Olsson, Leif A. Eriksson, and Johan Gottfries. Manuscript in preparation

# CONTRIBUTION REPORT

For Paper I, II, III, I was involved in planning the project and updating the progress of the project with Morten Grotli and Leif Eriksson. I was responsible in designing, performing calculations, analysing data and writing the paper.

For Paper IV, I was involved in planning the project with collaborators. I was responsible for performing the calculations, analysing data and writing parts of the paper.

For Paper V, VI, I was involved in planning and designing the project with Noel Lownds and Leif Eriksson. I performed theoretical calculations and analysed data. I was responsible for writing the paper.

Paper VII. I was involved in planning and designing the project with Leif Eriksson. I was responsible for performing calculations, analysing data and writing the paper.

# LIST OF ABBREVIATIONS

Abbreviations commonly used in this thesis:

ADME/T – Absorption, Distribution, Metabolism, Excretion and Toxicity (ADME/T)

AR – Androgen Receptor

CADD – Computer Aided Drug Design

DMPK – Drug Metabolism and Pharmacokinetics

DSB – DNA Double-Strand Break

EGFG – Epidermal Growth Factor Receptor

FMTC – Familial Medullary Thyroid Cancer

GDNF – Glial cell-derived Neurotrophic Factor

HAT – Histone Acetyltransferases

HTS – High-Throughput Screening

LBDD – Ligand Based Drug Design

MD – Molecular dynamics

MEN2 – Multiple Endocrine Neoplasia Type 2

PDB – Protein Database Bank

PTC – Papillary Thyroid Cancer

QSAR – Quality Structure Activity Relationship

RET – Rearranged during Transfection

RMSD – Root Mean Square Deviation

RTK – Receptor Tyrosine Kinase

SBDD – Structure Based Drug Design

# LIST OF CONTENTS

# Chapter 1.  An overview of drug discovery and development

In this chapter, I will briefly introduce some advances in drug discovery area, focusing on progress in cancer and cardiovascular therapeutic agents; following on with the drug discovery process and the current status of developing anticancer agents.

## 1.1 Drug discovery advances

Our body is made up of various types of cells, which carry out complicated molecular reactions to perform different functions, such as digesting, moving or thinking. In the cells, one type of molecule (gene or protein) interacts with another which, in turn, affects another, and so on in order to initiate or regulate the expression of specific proteins. These cascades of molecular interactions/reactions are called signalling pathways. However, if mistakes occurred in a signalling pathway, i.e. mutation, it can lead to that production of an important protein is halted, or overexpression. For example, these molecular disorders result in extra cells to grow in cancer, or cause our body to not produce enough insulin in diabetes. Drug molecules are able to affect the disordered pathways by interacting with certain molecules involved in the pathway, and either restoring normal activity or hindering e.g. tumor cell growth, in order to achieve the purpose of treatment.

Over the course of the last two centuries, diseases and conditions that were once considered as incurable or fatal have been conquered with therapeutic medicines designed to extend and improve quality of life. Modern drug discovery and development has clearly changed the course of disease treatment. For example, a significant portion of the improved clinical outcomes in cancer can be attributed to the discovery and development of novel therapeutic agents. The discovery of natural products and natural pro10duct analogues of antitumor activity such as Paclitaxel[1], Vinblastine[2], Doxorubicin[3], and Topotecan[4], together with the development of small molecule kinase inhibitors such as Imatinib[5], Nilotiniband[6] and Erlotinib[7] provide apparent evidence of the power of modern drug discovery in the cancer treatment. Meanwhile, the discovery and development of novel therapeutic agents have dramatically make survival rates better for many types of cancer. In the United States, from 1950 to 2009, the overall cancer death rates have declined 11.4%, and progress against some specific cancer types has been considerable. During the same period, the 5-year survival rates for breast cancer, prostate cancer, and melanoma have been significantly improved: breast

cancer increased from 60% to 91%, prostate cancer increased from 43% to over 99%, and melanoma increased from 49% to 93%[8].

Numerous therapeutic agents have also been designed to mitigate symptoms or prevent the underlying causes of the cardiovascular disease. For example, Amiloride[9], Indapamide[10], Atenolol[11], Propranolol[12], and Captopril[13] are just a few types of compounds currently available to lower blood pressure and improve in both the quality of life and life expectancy of patients. Further, the HMG-CoA reductase inhibitors, also known as statins[14], such as Atorvastatin[15], Simvastatin[15] and a number of related agents have shown remarkable capacity to lower cholesterol levels, which is a major risk factor associated with cardiovascular disease[16]. Similar improvements in the treatment of infectious disease, pain management, respiratory disease, and many other conditions has been described as well, due to profound and positive impact brought by the identification of novel therapeutic agents. In general, there is no doubt that drug discovery and development brings about longer and increased quality of life for individual and has positive impact on society.

**1.2 Drug discovery process**

Drug discovery and development is very lengthy and complicated, the whole process is time and resource consuming, and requires collaborations from a wide array of expertise in many fields, such as medicinal chemistry, drug metabolism, animal pharmacology, process chemistry, clinical research, etc. Further, high throughput screening, combinatorial chemistry and molecular modelling also play important roles in modern drug research. At present the cost invested in the drug discovery and development of a new drug ranges from $800 million to $1.8 billion, and the time of bringing a new drug to market, normally takes about 7–12 years[17]. Furthermore, it has been indicated that to identify a single marketed drug, it needs an initial screening of over 100,000 candidate compounds, hundreds of preclinical animal testing, and various clinical trials involving thousands of volunteers and patients. A recent report of clinical trial success rates has shown that only 1 out of every 10 compounds of clinical candidates will successfully pass through clinical trials and reach the market. This represents a success rate of less than 0.001% if measured by the number of compounds tested at the beginning of the process[18]. The above data is just a estimation, but illustrates the average time, money, human and other resources involved in developing a new drug.

The process by which a new drug from identified to marketed is referred as the development chain or "pipeline" and consists of a number of distinct stages (Figure 1.1). It can be divided

into two major stages: drug discovery and drug development. The first stage, drug discovery, can be further divided into three different steps: target discovery, lead discovery, and lead optimization. In this stage, a series of experiments and studies are designed to carry out initial identification of a biological target, as well as search for a single compound to control the activity thereof. Once a single compound has been identified, it goes to the second stage, which is drug development. The compound is then progressed through numerous studies designed to validate its safety and efficacy and support its approval for sale by the appropriate regulatory bodies.

The first step, target identification, is to identify the biological target within the body (the protein) and design a model to mimic human disease state. This step is typically fulfilled through the use of molecular probes, which can identify multiple series of compounds that have the ability to regulate the activity of the biological target of interest. The second step, lead discovery, is to identify a lead compound (molecule), which will exhibit drug-like properties. The third step, lead optimization, is to optimize the lead compound with respect to target protein of interest so that it may enter into the drug development phase. Subsequently, the drug goes through the preclinical phase of animal pharmacology and toxicology studies, as well as formulation, stability studies, and quality control measures. Finally, if the drug passed all the steps above, the drug enters to many phases of clinical trials in humans. In the clinical trials, the drug is administered to human volunteers,

- Around 20-80 healthy participants, the purpose of Phase I is to determine the maximum tolerated dose of the drug, to look at how the body handles the drug and to check for adverse effects

- Around 100-300 participants, the purpose of Phase II is to ascertain that the drug candidate actually has the desired effect on the illness, to identify optimal dose for drug use in humans and to identify ineffective medicines at an early stage

- Around 1000-3000 participants, the purpose of Phase III is to insure that proposed drug is safe, i.e. does not have frequent or severe side-effects, and to make sure it is effective and has advantages over existing drugs targeting the same illness

Only if the above criteria have been accomplished, approval for general use and marketing as a drug might be obtained from the appropriate national regulatory agencies (FDA, the Food and Drug Administration, in the US and the EMA, European Medicines Agency, in the EU).
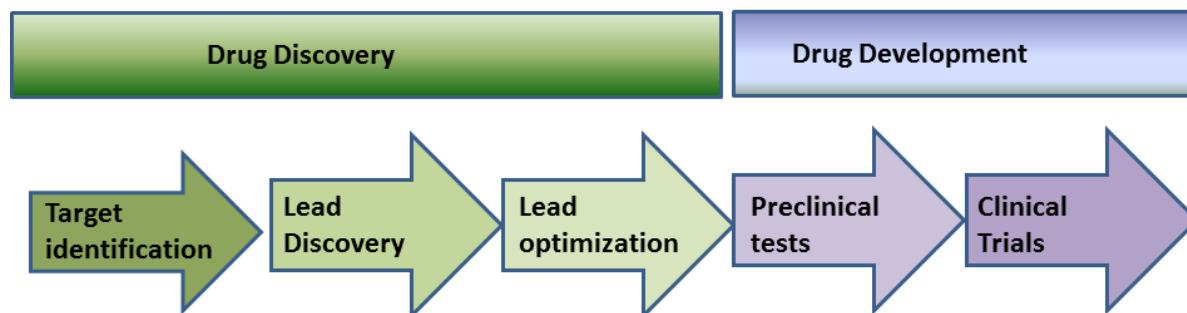
**Figure 1.1:** The drug discovery and development process

## 1.3 Cancer drug discovery

Nowadays, it is estimated that there are over 200 different types of cancer, depending on which part of body where they first are recognized. Cancer can happen at any age, but it is much more common in people over 65 years old. Therefore, in the next 20 years, it is likely that we will see a large increase in the incidence of cancer due to the aging of the world's population. The most common human cancers occur in lung, bowel, prostate and breast, and the less common cancers occur in blood and lymphatic systems, as well as the brain[19].

Normal cell growth is very highly ordered and carefully regulated by signals that dictate whether the cell should divide and grow into two cells or not, so the cells only grow when they are required. Cancer cells, however, are a group of abnormal cells, which grow uncontrollably by ignoring the normal rules of cell division. In this way, cancer cells develop a degree of independency from these signals, leading to uncontrolled growth and development into a mass or tumor. If this kind of proliferation is allowed to continue and spread to other parts of the body, which is a process called metastasis, it can be fatal. In fact, almost 90% of cancer-related deaths are due to metastasis[20].

Surgery, radiotherapy and drug therapy are the three basic ways to treat cancer, and can be used alone or in combination to eliminate discrete tumours locally. Drugs, as opposed to surgery and radiotherapy, have the ability to reach practically every part of the body and act systemically. The advantage of cancer drugs is that they can be used to eliminate the tumors locally, meanwhile, they are also possible to kill any individual cancer cells that are detached from the main tumor and tried to spread to other parts of the body, even when the spread is difficult for radiological or laboratory tests to notice yet. Furthermore, cancer drugs, may, on the one hand be used as main form of treatment to reduce the tumor size and alleviating symptoms. On the other hand, they can also be used before surgery to reduce the tumor size

and make it easier to remove, or after surgery and radiotherapy to reduce the chance of reoccurrence.

There are four broad classes of cancer drugs as follows:

**1,** Cytotoxic drugs. These drugs predominantly kill growing and dividing cells by influencing a cell's ability to divide. Since they do not specifically target cancer cells, they give rise to a high incidence of side effects. The most notable examples include infection, fatigue, hair loss, and severe nausea and vomiting, as the cells of the immune system, the blood, the hair follicles and the gut are also affected by the drugs. Furthermore, cytotoxic drugs have little effect on other aspects of tumor progression such as tissue invasion and metastasis.

**2,** Endocrine therapy. The drugs that belong to endocrine therapy are sex hormones, or hormone-like drugs, so they are basically used to change the action or production of hormones in the body. Some tumours have been found to grow in response to natural hormones. For example, certain breast cancers are stimulated by normal female hormone oestrogen, and prostate cancer is initially provoked by the male hormone testosterone for its growth. Endocrine therapy acts to make the cancer cells unable to utilize the hormones needed for growth, or prevent the body from producing such hormones.

**3,** Targeted therapies. The aim of targeted therapies is to specifically act on a well-defined target or biologic signalling pathway. Cancer cells and normal cells are differentiated in terms of their genes or the proteins in the cells. Targeted therapies make use of these differences by targeting the specific factors that are different in the cancer cell. Thereby targeted therapy can in principle eliminate cancer cells without disturbing the functions of normal cells and tissues, by inhibiting the cancer progression driven by the molecular pathways that transmit the pathological signals. Targeted drugs can be categorized into two classes, namely small molecule drugs and large proteins. Small-molecule drugs usually have the ability to enter the cell to block specific pathways, thereby preventing cell proliferation and resulting in cell dysfunction and death. They normally inhibit certain enzymes in cancer cells, such as the tyrosine kinase receptor. Large proteins, such as monoclonal antibodies, are able to attach themselves to receptors on the surface of the cancer cells and prevent signals being transmitted into the cell. In this thesis, we will focus exclusively on small molecule targeting.

**4,** Vaccines. Therapeutic vaccines are developed to treat cancers. These are given to people with cancer to strengthen the body's own immune system and stimulate it to recognize and attack the cancer cells. These types of vaccines may thus inhibit the further growth of existing cancers, prevent recurrence of treated cancers or kill cancer cells that already exist. Vaccines

are being tested in different cancers, such a breast, colorectal, lung and renal cancers and in B-cell non-Hodgkin's lymphoma, however, no vaccines have as yet reached the market. Acknowledging that drug discovery is difficult and complicated in general, oncology has one of the poorest records for drugs to reach into clinical trial phases, with success rates e.g. three times lower compared to cardiovascular disease[21, 22]. The overall success rate for anticancer agents in clinical development is about 10%, and the approval rate of new compounds for cancer treatment has declined steadily over the past decade. To this end, situation of anti-cancer drug discovery and development is currently less optimistic and progressive. Up until the 1990s, anticancer drug development was focused on the identification of cytotoxic compounds, which demonstrated apparent cytostatic or cytotoxic activity on tumor cell lines and leading to tumor regression. Since then, the developments in molecular biology and increased understanding of cancer at a molecular level, have been driving anti-cancer drug development move into the ''molecular target'' era [23, 24]. Molecular targeting in the treatment of cancer is based on the precondition that the cancer cells contain molecular marker (normally enzyme) considered to be important in cancer prognosis, growth, and/or metastasis. This molecular marker in cancer cells is normally caused by genetic abnormalities, resulting in increased and/or decreased production or activation. By determining the structure of the target enzyme, it allows for the design of selective and effective inhibitors targeting that specific enzyme. This enables selective inhibition of the characteristic molecular signalling pathway in cancer cells, with low effect on normal cells and, therefore, low toxicity. Some successes have emerged as a result of this strategy in recent years, such as imtatinib, which is a small molecule that inhibits a specific tyrosine kinase enzyme of the Bcr-Abl oncoprotein, and has been used for gastrointestinal stromal tumor and chronic myeloid leukemia. Gefitinib, another small molecule inhibitor for the tyrosine kinase enzyme - epidermal growth factor receptor (EGFR), has been used for non-small-cell lung cancer (NSCLC). Therefore, the molecular target strategy has led to the expectation of a new ''golden era'' in cancer drug development[25]. However, there are also some problems associated with this strategy. For example, inhibiting a few targets might not be sufficient, as progression of a normal cell to cancer cell has been shown to involve dozens of genetic mutations or signalling pathways. Furthermore, inhibition of a single molecule target with high selectivity can lead to resistance development to the cancer drug[22]. Although there are problems with molecular targeting, the underlying principles of this strategy are clearly sound, and we should use it as one part of a broad approach to the treatment of cancer and use it in conjunction with other treatments. In

this thesis, based on the molecular targeting strategy, we aim to develop anticancer drugs by designing or searching for inhibitors for varies types of enzymes or receptors.

# Chapter 2.   Computer aided drug design (CADD)

Drug discovery and development is a lengthy process which includes searching for promising hits, translating hits to leads, and final validating leads to drug candidates in clinical trials. Over the past decades, the investment in new drug development has increased considerably. However, despite considerable efforts the output is hampered by the low efficiency and high failure in the drug discovery and development process. Computer aided drug design (CADD) is one of the most effective new methods for facilitating and expediting this process, and therefore save time, money and resources[26].

## 2.1 History progress of CADD

The start of intense interest in CADD dates back to 1981, when Fortune magazine entitled a cover article as "Next Industrial Revolution: Designing Drugs by Computer at Merck"[27, 28]. Back to 1970s, CADD had already been established to regulate the biological activity of insulin and to guide the synthesis of human haemoglobin ligands based on the existed available knowledge of structural biology. Over the years, with high-throughput screening (HTS) technologies emerging, the drug discovery and development process was expedited and facilitated by enabling a great amount of compounds to be screened in much shorter period of time. HTS is a method especially used in drug discovery to quickly conduct millions of assays relying on automation, in order to search for compounds that evoke the desired biologic response. However, the hit rates are often extremely low, which limits the usage of high-throughput screening (HTS) techniques. Further, even though the hit is identified, it is still possible to fail in the following stage, such as optimization to lead or preclinical/clinical test, and the late stage failure was around 40–60% due to absorption, distribution, metabolism, excretion and toxicity (ADME/T) shortage. Therefore, all these issues underline the need to develop alternative strategies that can help in promoting the success rates and reducing the cost and time in the whole drug discovery and development process. These, together with advances in X-ray crystallography, HTS techniques and computational power in 1990s, have rapidly improved the progress of CADD application in pharmaceutical industry. Numerous of approved drugs that credited their discovery in large part to the tools of CADD were reported, such as carbonic anhydrase inhibitor Dorzolamide for the treatment of cystoid macular edema[29]; the angiotensin-converting enzyme (ACE) inhibitor Captopril for the treatment of hypertensive[30], human immunodeficiency virus (HIV)

protease inhibitors Saquinavir, Ritonavir, and Indinavir for the treatment of HIV[27]; and renin inhibitor Aliskiren, which is used for essential hypertension[31].

Nowadays, CADD is playing an increasingly larger and more important role in drug discovery and development and helps improving efficiency for the industry. One of the most frequently used tool in CADD is the screening of virtual compound libraries, also termed virtual screening. Compared to traditional HTS, which requires extensive preparation and lacks the primary understanding of the molecular mechanism behind the activity of identified hits, virtual screening requires significantly less workload and uses a much more targeted search based on known ligand or target structures. In 2003, a group at Biogen Idec used virtual screening to search for inhibitor to target transforming growth factor β1 receptor kinase[32], and another group in Eli Lilly used a traditional HTS for the same target[33]. The group in Biogen Idec identified 87 hits of which one shared identical structure to the lead discovered in Eli Lilly by traditional HTS. This example demonstrated that virtual screening in CADD is able to produce the same compounds as a full HTS procedure, however, with significant less cost and workload. Over the years, a large fraction of hits has been found for kinases and G-protein-coupled receptors (GPCRs) by virtual screening[34]. It was concluded that "The future is bright. The future is virtual".[35]


## 2.2 CADD applications in drug discovery and development

CADD tools have been applied in almost every stage, greatly changing the strategy and pipeline for drug discovery (Figure 2.1). Although the traditional application of CADD is in lead discovery and optimization, today the application extends in the direction of target identification and validation, and forwards preclinical studies, mostly through ADME/T prediction. In the drug discovery and development process, CADD is usually used for three major purposes: (1) screen large compound libraries into smaller sets of compounds, in order to experimentally test only the compounds with highest predicted activities; (2) instruct the optimization of lead compounds, in order to increase the binding affinity or optimize drug metabolism and pharmacokinetic (DMPK) properties including ADME/T; (3) design novel compounds, either by "adding/modifying" functional groups in starting molecule or by linking together fragments into novel compounds[17, 28, 36].

An example of applying CADD in different stages of the drug discovery process is illustrated in Figure 2.2. We start from applying homology modelling to generate a target structure, followed by molecular dynamics (MD) simulations to optimize the target structure. Then the protein structure is ready for docking of compound libraries to identify potential binders.

Once the possible binders are identified, combinatorial chemistry can be used to generate a series of derivatives. However, if there is no target structure available, a QSAR pharmacophore can be generated based on ligand structure and activity information, where key pharmacophore features can be achieved for searching the same classes of binders to the target. Further, the DMPK properties of the binders, such as ADME/T can also be predicted by CADD tools and used to compare with bio-assay data. If a compound can pass all the steps above, then it becomes a drug candidate for the following clinical trials.
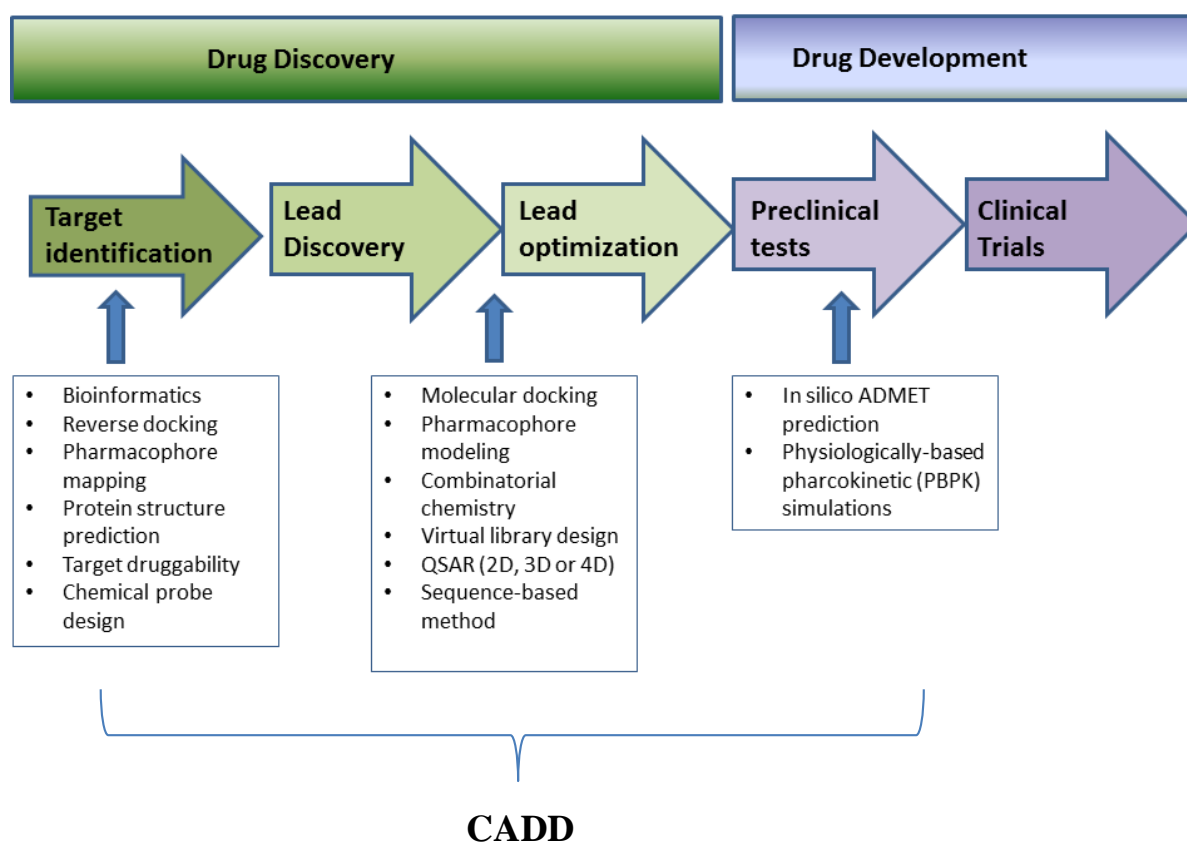


**Figure 2.1:** Multiple computational drug discovery approaches have been applied in various stages of the drug discovery and development pipeline, including target identification, lead discovery and optimization, and preclinical tests.

## 2.3 Classification of CADD

CADD can be categorized into structure-based and ligand-based[37]. Structure-based drug design (SBDD) relies on the availability of a 3D structure of the biological target obtained through methods such as X-ray crystallography or NMR spectroscopy. If no experimental

structure of a target exists, it may be possible to generate a homology model thereof based on known 3D structures of related proteins. Based on the structure of the biological target, docking can be applied to place each ligand, typically a molecule or molecular fragment, into the binding site of the target, and predict its most favourable binding mode by ranking the predicted activity of each compound in terms of the estimated binding affinity. Moreover, the essential interactions between the ligand and the binding site of the receptor can be translated into a structure-based pharmacophore model which can be used to screen a large database for possible active ligands. Both docking and structure-based pharmacophore model screening is about finding ligands for a given receptor; in both cases, large libraries of compounds are screened to find those fitting the binding pocket of the receptor. Another method, de novo design, involves directly building ligands within the constraints of the binding pocket by adding small pieces, either individual atoms or molecular fragments, in a stepwise manner. In addition to the above techniques, MD simulation also requires a comprehensive understanding of the target structure, and is therefore also included in SBDD category.

If the 3D structure of the target, the binding site or even the target itself are not accurately known, then Ligand-based drug design (LBDD) is an appropriate method to apply if there are experimentally active compounds that bind to the biological target of interest. These compounds may be used to derive a ligand-based pharmacophore model, in which the minimum necessary structural features a molecule must contain in order to bind to the target are defined. In addition, a model of the biological target may be generated based on the known compounds that interact with it, and this model may in turn be used to search for new molecular entities with the same features that can then be expected to bind to the target, This is termed ligand-based virtual screening. Furthermore, quantitative structure-activity relationship (QSAR) can be used to derive a correlation between theoretically calculated properties of molecules and their experimentally obtained biological activity. The resulting correlation derived from QSAR can in turn be further used to predict the activity of new analogues.

## 2.4 Limitations of CADD and future outlook

CADD has already provided significant benefits, and holds a great promise for future progress in drug discovery and development. However, it is still an evolving technology and has a number of limitations[38-40].

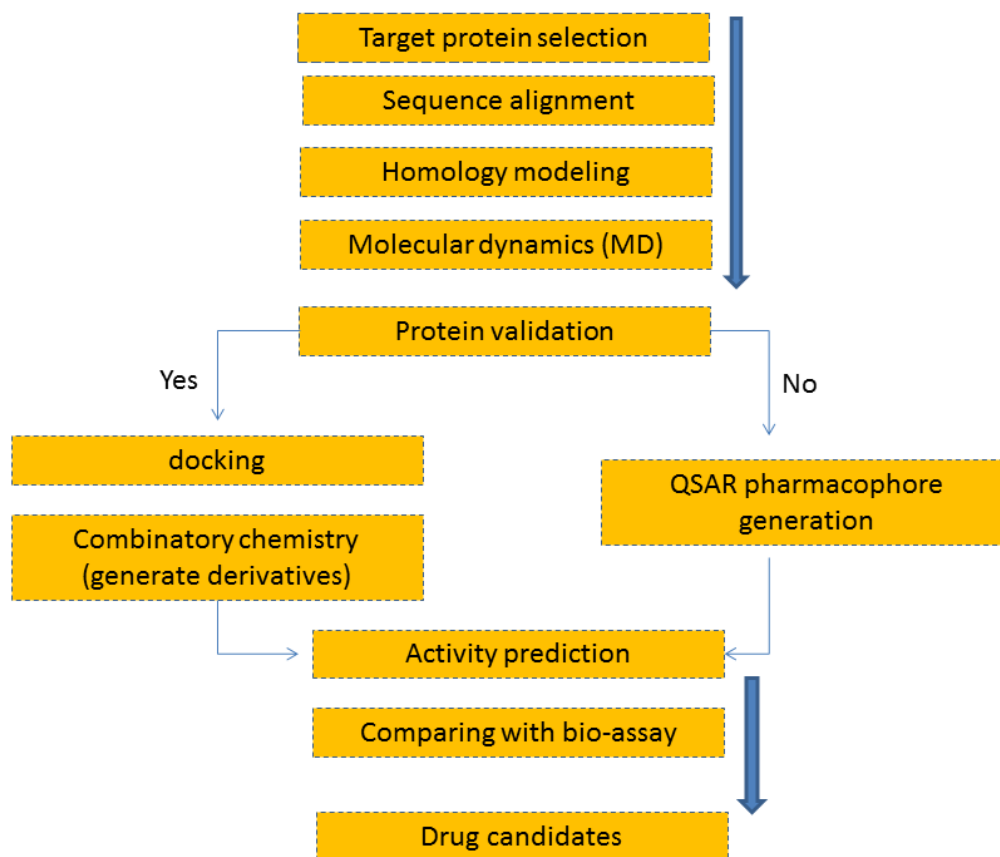**Figure 2.2:** An example of workflow by using CADD in drug discovery and development process
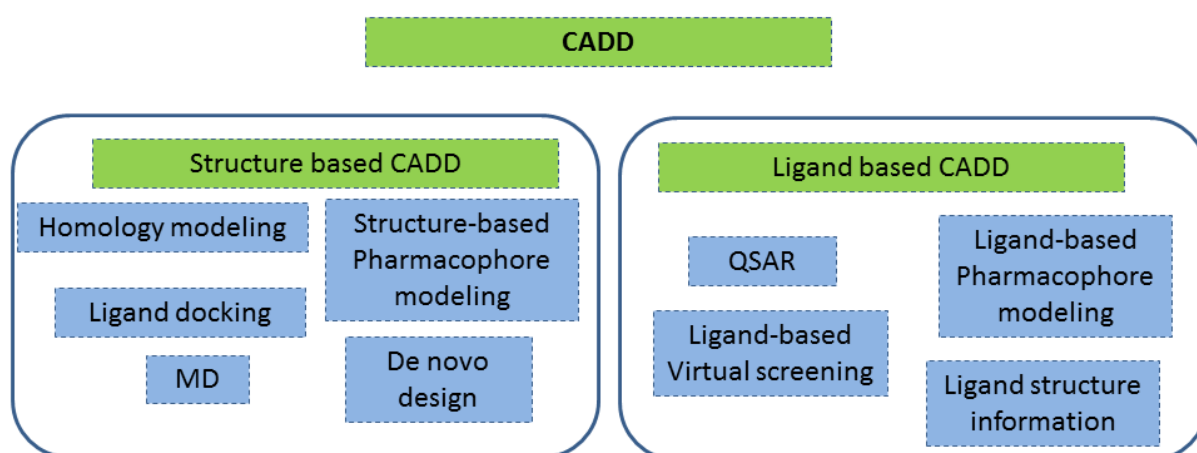


**Figure 2.3:** CADD is classified into two groups based on the availability of target structure information.

In pharmacophore modelling, for example, due to the induced fit effect, a receptor may adapt to different ligands. Therefore, it is possible to have multiple pharmacophores for a single site.

Obviously, generating a common pharmacophore from a set of ligands containing different scaffold is not an optimal choice[41]. Docking also has its limitation[42, 43], for example, when increasing the molecular mass and number of rotatable bonds of the compounds, the number of possible conformations also increase, which presents intense demands on computational hardware and software; binding be coupled with an induced fit effect that leads to protein adaptability; however, these additional conformational changes are normally not considered; the importance of the crystallization process and how representative a single crystal structure is, is often neglected[44]; the role of solvent molecules is difficult to verify, although in many cases, solvent molecules are important by providing bridging hydrogen bonds for the ligand to interact with binding site. It is reported that docking programs currently are able to dock 70–80% of ligands correctly into the binding site[45]. One recent study proposed that the significant issue in structure-based virtual screening is mostly due to the current docking and scoring algorithms are unable to identify key interactions and treat them appropriately[43].

In the future, the primary task is to increase the accuracy and effectiveness of existing technologies in CADD[46]. In addition, it is also important to integrate computational chemistry and biology together with cheminformatics and bioinformatics in CADD, in order to exploit or apply all the available knowledge from different subjects. The completion of the human genome and numerous pathogen genomes could be of great help for discovering new drug targets, hence more focus will undoubtedly come on computational methods involving target identification. Tailored small molecules will also be extensively applied as probes for functional research.

# Chapter 3.   Anti-cancer target in this thesis

In this chapter, several protein targets will be described, which have been shown to relate in cancer development. The signalling pathway, the mechanism behind the induced cancers and the current available inhibitors for the targets will be introduced in detail, from which we will understand the significance of investigating these targets.

## 3.1 The tyrosine kinase RET

RET (rearranged during transfection) encodes a transmembrane receptor tyrosine kinase (RTK) and is essential for development of the peripheral nervous system, kidney morphogenesis and spermatogenesis[47]. RET protein is composed of three domains, an extracellular domain, which contains the cadherin-like domain (CLD) and cysteine-rich domain (CRD), a transmembrane domain (TD), and an intracellular portion containing the tyrosine kinase domain (TKD). RET is activated through binding to a soluble, bivalent GDNF family of ligand (GFL), which is comprised of glial cell-derived neurotrophic factor (GDNF), neurturin (NRTN), artemin (ARTN) and persephin (PSPN), in complex with a preferred GPI-linked RET co-receptor (GFRα1-GFRα4)[48]. The formation of the RET/GFL/GFRα complex results in RET homodimerization and triggers autophosphorylation of intracellular tyrosine residues. Phosphorylated residues including tyrosine 687 (Y687), serine 696 (S696), Y752, Y791, Y806, Y809, Y826, Y864, Y900, Y905, Y928, Y952, Y981, Y1015, Y1029, Y1062, Y1090, and Y1096 constitute docking sites for numerous downstream signalling effectors, which then activate the signalling pathways, including the mitogen-activated protein kinase pathway (Ras/RAF/ERK)[49], the phosphatidylinositol 3-kinase/protein kinase B pathway (PI3K/AKT)[50], the c-Jun N-terminal kinase pathway (JNK)[51], and the signal transducer and activator of transcription 3 (STAT3)[52]. More recently, it has also been shown that RET carrys out direct tyrosine phosphorylation of beta-catenin, which is associated with an induction of RET tumorigenic ability *in vivo* [53] (Figure 3.1). Many of the above mentioned intracellular signalling pathways are activated not only by RET, but also by other RTKs. In general, the activated downstream signalling pathways contribute to the further regulation of cell survival, differentiation, proliferation and migration.

Gain-of-function mutations of RET are related to multiple endocrine neoplasia type 2 (MEN2), and papillary thyroid cancer (PTC)[52]. MEN2 is a hereditary cancer syndrome and it has three subtypes based on the clinical symptoms: MEN2A, MEN2B and familial medullary
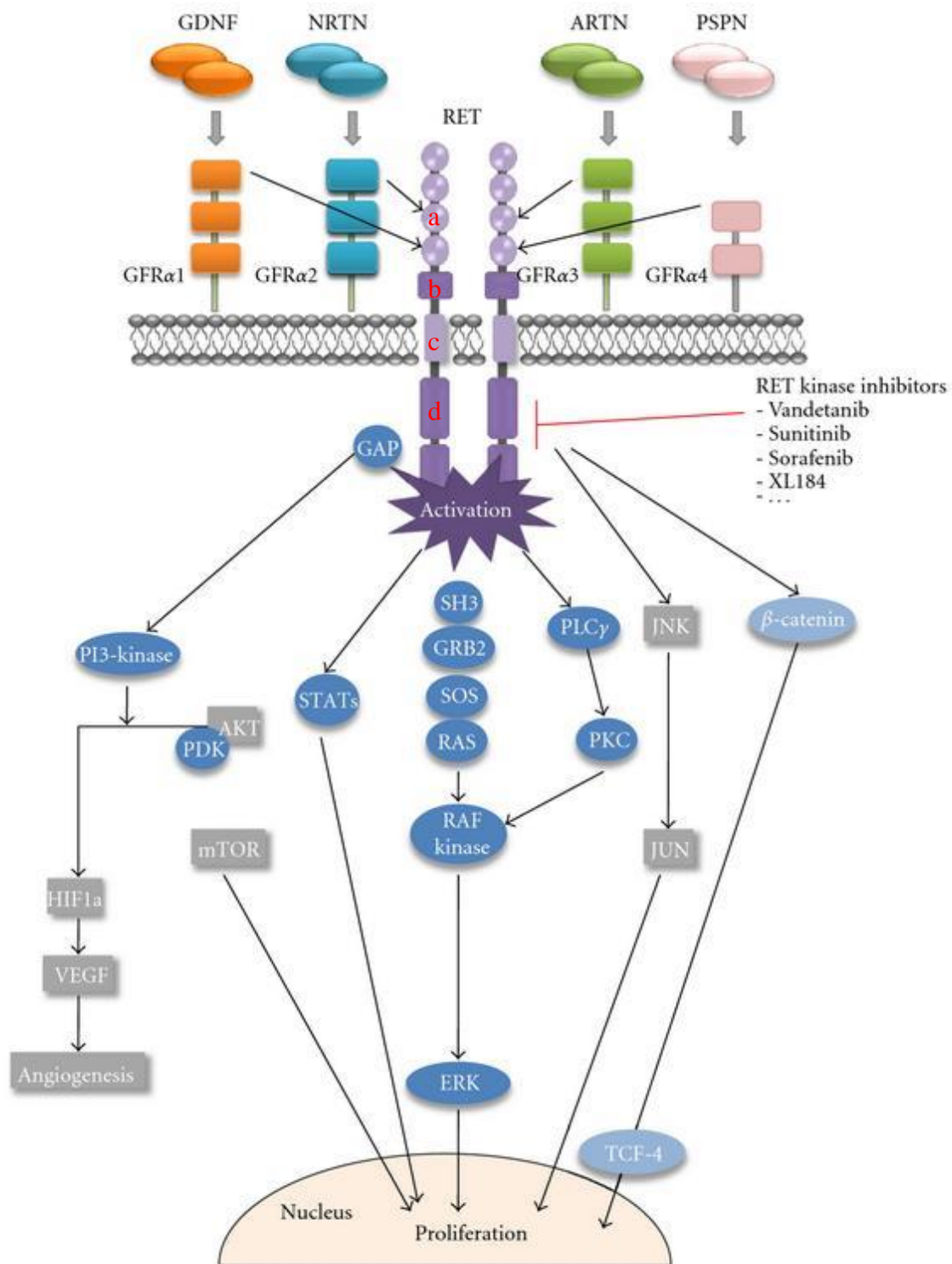
**Figure 3.1:** Outline of RET signalling pathway [54]. a, CLD; b, CRD; c, TD; d, TD.

thyroid cancer (FMTC). The molecular mechanism of RET activation in human cancer varies, at the germline level, point mutations of RET are responsible for MEN2. Mutations of extracellular cysteines at codons 609, 611, 618, 620, 630, predominantly, 634, are found in

MEN2A patients, and mutation at codon 918 of Met918 to Thr918 is responsible for most MEN2B cases. FMTC mutations are similar to those causing MEN2A, but are more evenly distributed among cysteines 609, 618, and 620. Moreover, in FMTC patients, mutations of residues 768, 790, 791, 804, 844, or 891 of the RET tyrosine kinase domain have also been found. At the somatic level, gene rearrangements cause the tyrosine kinase domain of RET juxtaposed to heterologous gene partners and lead to the formation of chimeric RET/PTC oncogenes, which are commonly found in PTC. Both MEN2 mutations and PTC gene rearrangements increase the likelihood of intrinsic tyrosine kinase activity of RET and RET downstream signalling events, with resulting cancer cell proliferation and metastasis.

Currently there is no available therapeutic option for treating RET associated cancer, although numerous kinds of therapeutic approaches, such as small molecules acting as tyrosine kinase inhibitors, gene therapy with adenoviral vectors expressing dominant negative RET mutants, monoclonal antibodies capable of internalization of RET have been developed[55]. The application of these strategies in preclinical models has demonstrated that RET is indeed a prospective target for selective cancer therapy. Of all those therapeutic ways to block the tyrosine kinase function of RET, small organic compounds show their potential for the treatment of human cancers possessing oncogenic RET. Moreover, those compounds also offer the possibility of interventional therapy, when conventional pharmacologic and radiotherapeutic regimens have failed. Several tyrosine kinase inhibitors including STI571, genistein, allyl-geldanamycin, and arylidene, RPI-1[56-58] are able to selectively inhibit RET tyrosine kinase activity and tumor cell growth in vitro. A combination of STI571 and PD173074, which is a fibroblast growth factor receptor (FGFR) restrains MTC cell growth by inhibiting both RET and FGFR[59]. CEP-701 and CEP-751, which are indolocarbazole derivatives, can inhibit MEN2A tumor growth in MTC cell xenografts[60]. PP1, a pyrazolo-pyrimidine derivative, blocks tumorigenesis induced by RET/PTC oncogenes, and induces degradation of activated membrane-bound RET receptors [61, 62]. Another pyrazolo-pyrimidine, PP2, also inhibits oncogenic RET activity[62]. The 4-anilinoquinazoline ZD6474 shows dual antitumor activity with a strong inhibitory activity towards constitutively active oncogenic RET kinases and angiogenesis[63].


### 3.2 The Histone acetyltransferase Tip60

Chromatin is a complex structure of DNA and proteins found in eukaryotic cells. The basic repeating structural unit of chromatin is the nucleosome, comprised of 146 base pairs of DNA

wound around a histone octamer. Histones are positively charged proteins that strongly attach to and compact negatively charged DNA, and the histone octamer is made up of two copies of each core histone H2A, H2B, H3 and H4[64]. The nulceosomes, in turn, are arranged into several tiers of higher-order structures that allow for packaging the genome into the microscopic space of the eukaryotic nucleus.
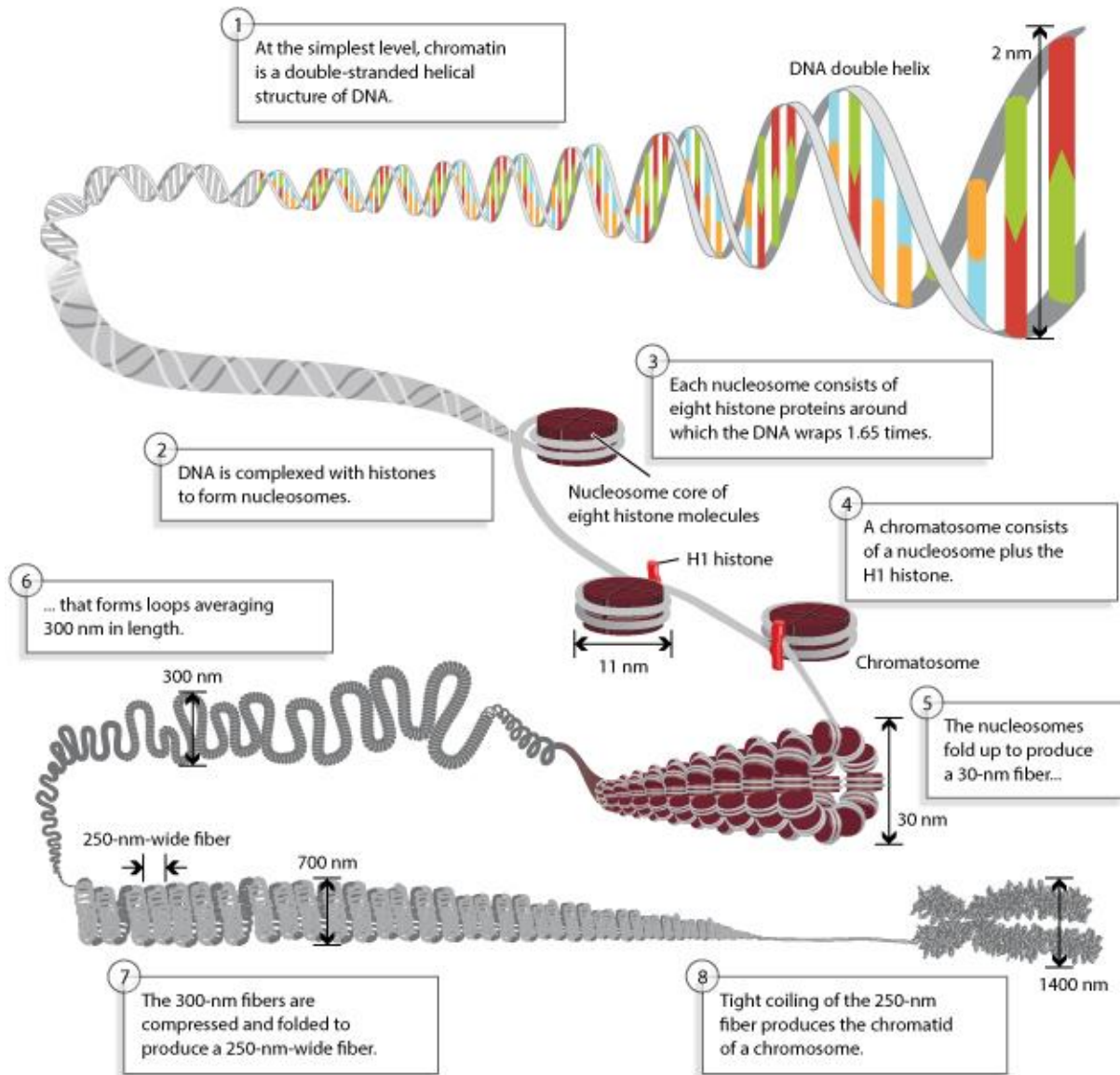
**Figure 3.2**: Chromosomes are composed of DNA tightly-wound around histones.

Chromosomal DNA is packaged inside microscopic nuclei with the help of histones. These are positively-charged proteins that strongly adhere to negatively-charged DNA and form complexes called nucleosomes[65].

At the cellular level, various DNA-based nuclear processes such as DNA transcription, replication and damage repair are continuously on going, therefore, in addition to its compactness, chromatin must also be very flexible and dynamic to permit those nuclear

processes. For this purpose, the local chromatin structure is required to be modified in order to make it accessible to non-chromatin proteins. Chromatin remodellers and histone modifiers are the two major classes of enzymes involved in such activities to assist chromatin flexibility[66]. Tip60 (also named KAT5) is one of the histone modifiers and belongs to the MYST family of histone acetyltransferases (HAT). There are currently five human HATs in this family, namely Tip60, MOZ, MORF, HBO1 and MOF. The defining feature of those five members is that they all contain a highly conserved MYST domain, which is composed of an acetyl-CoA binding motif and a zinc finger[67, 68]. Tip60 acytelation involves the transfer of an acetyl group from acetyl-CoA to the N-terminal of a lysine residue, thus changing the surface charge distribution of the histone and the accessibility to DNA and/or to other proteins[69].

Besides histones, Tip60 can also acetylate some transcription factors, including the androgen receptor (AR), myelocytomatosis oncogene c (c-Myc), upstream binding transcription factor (UBF) and the kinase Ataxia Telangiectasia mutated (ATM)[70]. Tip60 acetylates histones on genes of these transcription factors, resulting in promotion of their activity. Some studies have demonstrated that acetylation by Tip60 activates androgen receptor (AR). AR is a hormone-dependent transcription factor, and its over production is strongly related to the onset of prostate cancer[71]. It was also shown that Tip60 is able to progress prostate cancer cells to hormone independence and resistance to chemotherapy, and that nuclear Tip60 acts as a co-activator of AR in the absence of ligand[72]. Thus, upregulation of Tip60 may permit advanced prostate cancer cells to survive in a hormone-independent fashion. Some other studies have also demonstrated that upregulation of Tip60 is linked to promotion of epithelial tumorigenesis. In addition, Tip60 may play a role in the induction of adult T-cell leukaemia/lymphoma and other cancers involving c-Myc oncogene, which is a potent promoter of cellular growth and proliferation, and is often deregulated in a variety of human cancers[73].

In response to DNA double-strand break (DSB), Tip60 plays different roles[74, 75]. On the one hand, it acetylates the lysine 120 residue of p53, which is a tumor suppressor well known for inducing either cell growth arrest or apoptosis after DNA damage, and leads to p53-induced cell apoptosis[76]. On the other hand, it acetylates histone H4 at the DSB sites to "open" the chromatin structure DNA to DNA repair proteins, such as the ATM kinase protein. Tip60 also activates the ATM kinase directly by acetylating it in response to DNA damage, finally resulting in DSB repair[77] (Figure 3.3).
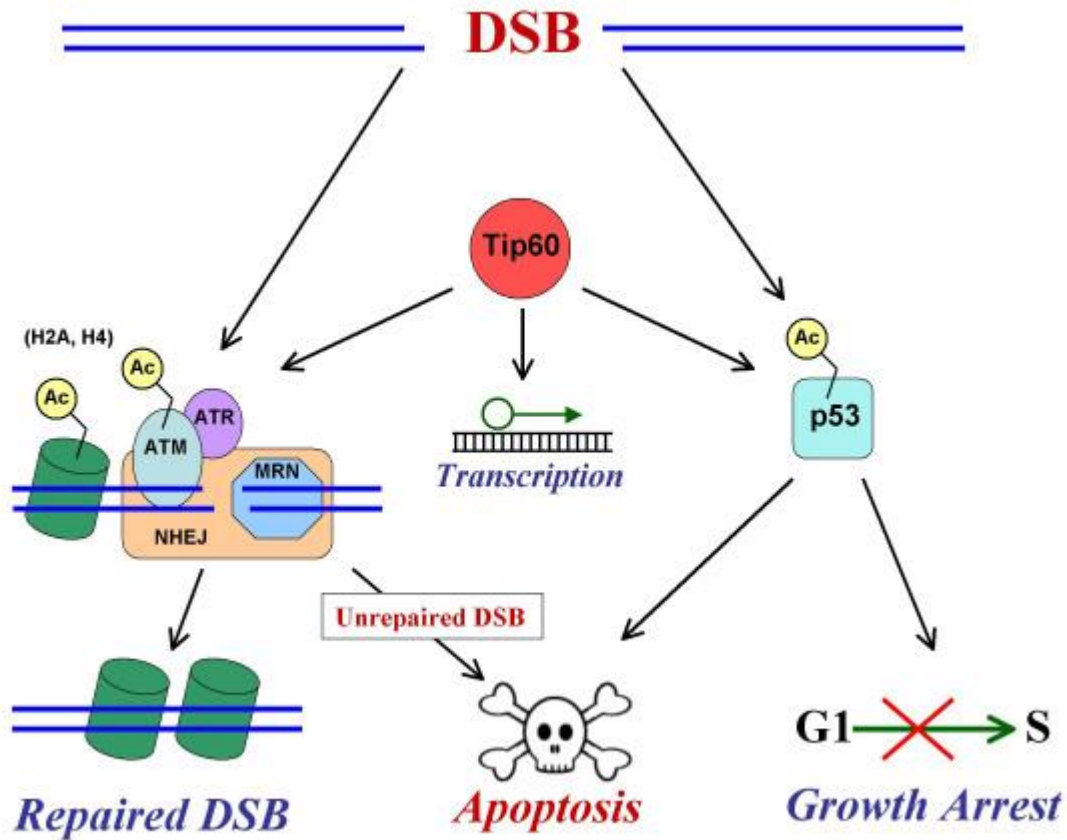
**Figure 3.3:** Model for the role of Tip60 in DSB repair[78].

Thus, based on the evidence that links Tip60 to key processes implicated in DNA damage response (DDR), DNA repair, cellular growth control and apoptosis, as well as its involvement of acetylation of transcription factors, it is possible to speculate that Tip60 might paradoxically function either as a promoter or as a suppressor of tumorigenesis. However, Tip60 inhibitor anacardic acid was shown to inhibit Tip60 *in vitro* and block the Tip60-dependent activation of the ATM protein kinase by DNA damage *in vivo*[79]. Further, it sensitizes human tumor cells to the cytotoxic effects of ioninzing radiation, which is a common treatment for more than half of human maglinancies, depending on its ability to damage DNA and kill tumor cells[80]. Therefore, Tip60 inhibitors could provide a novel therapeutic approach for increasing the sensitivity of tumors to radiation therapy. In addition, a small number of other HAT inhibitors have been reported today. Bisubstrate inhibitor, which is formed by coupling a histone H3 peptide to CoA, has been shown to decrease HAT activity; however, the compound has disadvantage of poor cell membrane permeability[81]. The natural product garcinol is HAT inhibitor that are cell permeable; and it sensitises cells to IR just like anacardic acid[82]. Isothiazolones, which are convalent binder interacting with the

28

HAT active site thiol, have been described as an effective starting point for further generation of more potent and specific inhibitors[83, 84]. Other HAT inhibitors include α-methylene butyrolactones, benzylidene acetones and alkylidene malonates[85-87].

## 3.3 Mitotic kinesins Eg5 and KIF18B

Mitosis is part of the cell cycle and is a fundamental process for cell division, in which replicated chromosomes undergo separation into two new nuclei by the mitotic spindle[88]. However, if a functional mitotic spindle does not form, normal chromosomal separation will not occur and checkpoint proteins will start to inhibit cell division, resulting in mitotic arrest[89]. The mitotic spindle consists of microtubule fibers that emerge from the spindle poles and attach to the condensed chromosomes at the centromere via the kinetochore. Microtubules are dynamic polymers constructed from α/β tubulin dimers. Currently drugs that target tubulin or microtubules are among the most effective cancer therapeutics, such as Vinca alkaloids[90], which inhibits tubulin polymerisation and thereby prevent mitotic spindle formation. Taxanes, on the other hand, stabilize GDP-bound tubulin in the microtubule and inhibit the spindle function by disrupting microtubule dynamics, which induces the mitotic arrest, followed by apotosis[91].

Though microtubules are essential for mitosis, they also participate in a number of other cellular functions, such as cell division, cell motility, intracellular transport and maintenance of organelles, synaptic vesicles and cell shape. Therefore, microtubule targeting drugs can lead to toxic side effects such as body weight loss, hair loss and neurotoxicity as seen with taxanes and vinca alkaloids[92]. Furthermore, carcinoma cells may become resistant to microtubule targeting drugs through various mechanisms, including mutations of tubulin, altered expression of tubulin subtypes, and overexpression of drug efflux pumps[93]. Therefore, there is a significant need to generate novel antimitotic drugs, with the aim to overcome the side effects and resistance seen with microtubule-targeting drugs.

Kinesins are motor proteins that move along microtubule filaments, and are important for many key cellular functions such as mitotic spindle assembly, microtubules remodelling, and chromosome separation in dividing cells. The kinesins are mainly composed of three regions, the "head" region, the "stalk" region and a "tail" region (Figure 3.4)[94]. The "head" region is the motor domain that contains an ATP binding pocket and a microtubule binding interface. The "stalk" and "tail" regions are needed not only for dimerization or oligomerization but also interaction with cargo. Kinesins use ATP hydrolysis to generate force and movement

along microtubules, and thus to enable microtubules to form the mitotic spindle and drive chromosome separation in mitosis[95].
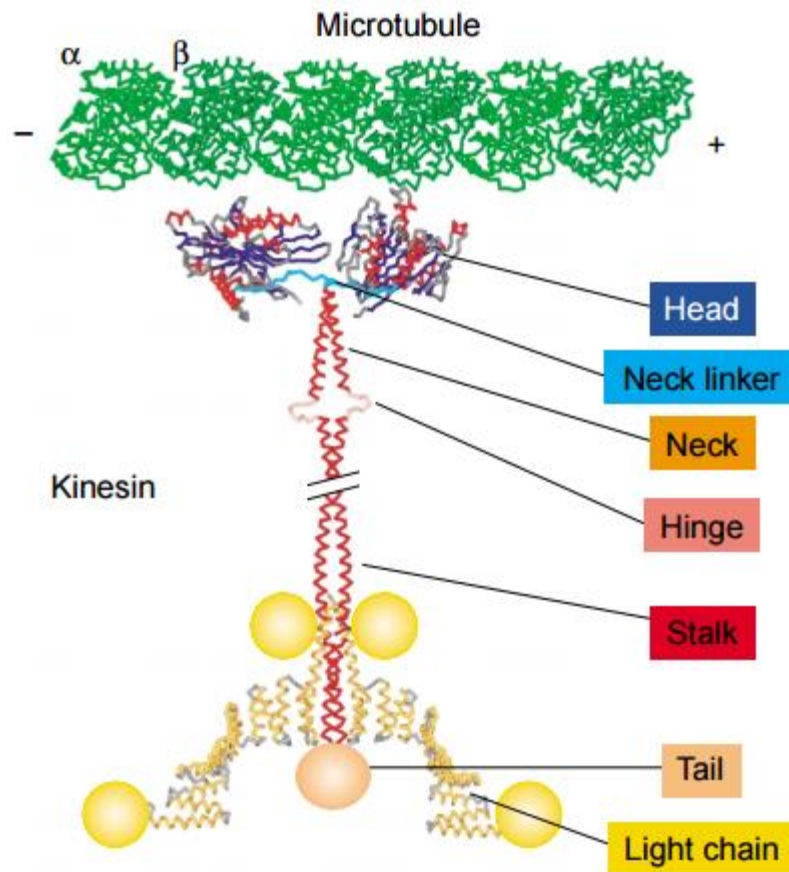


**Figure 3.4:** Kinesin domain structure[96]. The blue and red structure in the middle represents a dimeric kinesin heavy chain (KHC). Each heavy chain contains a motor domain ('head', α-helices red, β-strands blue) that binds to ATP and microtubules, a neck linker (cyan) whose conformation changes during the ATPase cycle, an α-helical neck and a stalk (red) that causes dimerization by a coiled-coil interaction. The stalk is interrupted by several non-helical hinges (only one shown here) that allow the heads to swivel and the stalk to bend over in a hairpin-like fashion, generating a 'folded' conformation that is inactive when kinesin is not moving. The tail binds to two light chains (KLC, ≈570 residues, yellow)

Eg5 is one of the mitotic kinesin family members that assemble the mitotic spindle during cell division (Figure 3.5). Inhibition of Eg5 prevents the mitotic spindle formation and centrosomal separation, leading to mitotic arrest without disturbing the microtubules. Therefore, Eg5 is considered as promising target for cancer treatment[98]. Compared to other

antimitotic inhibitors, inhibitors of Eg5 have more advantages[99]. First, Eg5 is found to be overexpressed in numerous proliferative tissues including leukemia as well as some solid
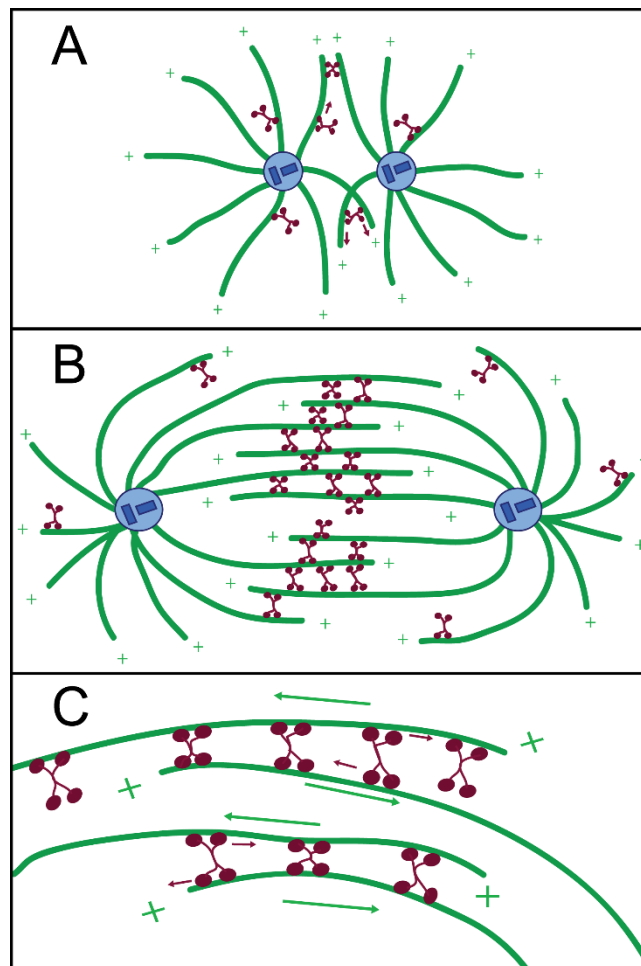


**Figure 3.5:** Schematic depicting Eg5 activity in the mitotic spindle[97]. Tetrameric Eg5 motors (red) help organize microtubules (green) to form the mitotic spindle. (A) At the onset of mitosis, the duplicated centrosomes (blue) separate and nucleate two microtubule asters. Processive Eg5 motors may translocate to the plus-ends of microtubules, located distal to the centrosomal organizing center and by crosslinking antiparallel microtubules, promoting bipolarity. (B) By metaphase, a stable bipolar spindle has formed. Eg5 motors likely provide structural integrity and also slide microtubules toward the centrosomes, contributing to the generation of poleward flux. (C) A close-up depiction of Eg5 motors walking to the plus ends of antiparallel microtubules, moving both poleward simultaneously.

tumors, such as breast, lung, bladder, pancreatic and ovarian cancers, however almost no Eg5 is detected in nonproliferative tissues[100-103]. Therefore, inhibitors of Eg5 arrest specifically cells in mitosis while not affecting any non-proliferating cells. Hence, inhibitors of Eg5 may

not have the severe side effects caused by traditional antimitotic inhibitors such as the Taxanes and Vinca alkaloids which target microtubules and affect both nonproliferating and proliferating cells. Second, Eg5 is not expressed in the peripheral nervous system of adults, and hence Eg5 inhibitors may not cause neuropathic side effects commonly found in inhibitors that primarily target tubulin. To date there have been seven Eg5 inhibitors introduced into Phase 1 or 2 clinical trials, and several more are in development[104]. The most known is SB-715992, which is the first Eg5 inhibitor introduced into human trials. Several other agents have since entered the clinic, including SB-743921, ARRY-520, MK-0731, AZD4877, EMD 534085 and LY2523355[105]. All of these compounds are ATP noncompetitive, and bind to the $\alpha2/L5/\alpha3$ allosteric pocket. Mutations in this binding pocket have been shown to cause resistance to SB-743921, providing a possible mechanism for drug resistance. Recent identification of several series of ATP competitive inhibitors that appear to bind in a region distinct from $\alpha2/L5/\alpha3$ pocket may provide a way to overcome this obstacle. Some compounds from these series, such as biphenyl type inhibitors, have been shown to overcome resistance and induce significant anti-tumour effects in mutant-Eg5 models, however reduced efficacy was noted with wild-type tumours in comparison with ATP uncompetitive allosteric inhibitors. Further development of selective ATP-competitive inhibitors that can be used in clinic, in combination with allosteric inhibitors targeting $\alpha2/L5/\alpha3$, could provide benefits in overcoming resistance that might arise from target mutation[99, 106].

KIF18B also functions as a mitotic kinesin, and it belongs to kinesin 8 family. loss of Kif18B results in an increase in the number and length of astral microtubules, suggesting Kif18B is an important modulator of astral microtubules dynamics[107, 108]. Further, KIF18B has also been demonstrated to involve in multiple tumors due to its deregulation in cell cycle[109], and by interacting with 53BP1, KIF18B is required for efficient double stand break repair[110]. Obviously, KIF18B function is important for mammalian organisms, however, how it exactly functions are still unclear as far as we know.


**3.5 GTPase K-Ras**

The Ras protein family members belong to low molecular weight GTPase (GTP-binding) proteins, involved in cellular signal transduction that transmit extracellular signals to intracellular effector pathways, and ultimately lead to cell differentiation, proliferation and survival[111]. H-Ras, N-Ras, and K-Ras are the three members of the Ras protein family, all of

which have been found to be related to cancer formation[112]. Mutations in Ras proto-oncogenes are frequently found, which is estimated to be 20-30%, in all human tumors. Making Ras mutations one of the most prevalent drivers of cancer[113, 114]. K-Ras is the most frequently mutated Ras member, having been shown to be mutated in 90% of pancreatic cancers, 45% of colorectal cancers, and 35% of lung cancers. In addition, K-Ras mutations have been associated with increased tumorigenicity and poor prognosis. The inhibition of activated Ras help malignant cells revert to a non-malignant phenotype and cause tumor regression both *in vitro* and *in vivo*[115]. Hence, K-Ras has become an attractive therapeutic target for various kinds of cancers.

K-Ras functions as a molecular switch that exchanges between GDP-bound and GTP-bound state (Figure 3.5). The GDP-bound form is generally considered switched off and inactive, while the GTP-bound form is switched on and active to stimulate the downstream pathways. The main conformational change of GTP- vs GDP-binding is located in the so-called switch I and II regions[116]. Generally Ras proteins bind to GDP proteins very tightly, and hydrolyse GTP at a very slow rate, so they need GTPase activating proteins (GAPs) to stimulate GTP hydrolysis and guanine nucleotide exchange factors (GEFs) to facilitate GDP dissociation. In the cellular context, the activated upstream signal is transmitted to Ras proteins, thereby stimulate the recruitment of GEFs such as SOS, which catalyze the exchange of GDP to GTP and switch on Ras proteins to promote downstream normal signalling, and normal cell growth or differentiation[117]. In the GTP bound form, Ras interacts with GAPs such as p120GAP, which increase the intrinsic activity of Ras proteins to hydrolyze GTP to GDP[117]. However, single point mutations of the Ras residues, such as mutated G12 and G13, abolish GAP-induced GTP hydrolysis through steric hindrance, while mutations of residue Q61 interfere with the coordination of a water molecule necessary for GTP hydrolysis[118]. These mutations lead to the constitutive active Ras proteins in active GTP bound form, thereby cause the constant activation of its downstream effector pathways, such as RAF-MEK-ERK, PI3K-AKT, and RALGDS-RAL-RLIP signalling pathway, which promotes oncogenic signalling and, ultimately cancer cell proliferation[117].

The K-Ras protein is considered a major target in anticancer drug discovery due to its high mutation rate in various cancers[115]. However, so far no clinically useful drugs directly interfering with K-Ras signalling have been found, because it is difficult and competitive to find inhibitors binding to K-Ras GTP binding pocket as the affinity of GTP for Ras is extremely high, which is in the picomolar range. Based on NMR fragment-based lead discovery and structure-based drug design, two groups recently managed to find novel small-

molecule inhibitors of K-Ras targeting an allosteric binding site[115, 119]. The allosteric inhibitors bind to a hydrophobic pocket between the Switch II and core β sheet region of K-Ras, with micromolar affinity. The allosteric site is different from but partially overlapping with the GEFs binding site such that GEFs are unable to activate K-Ras when the inhibitors bind.
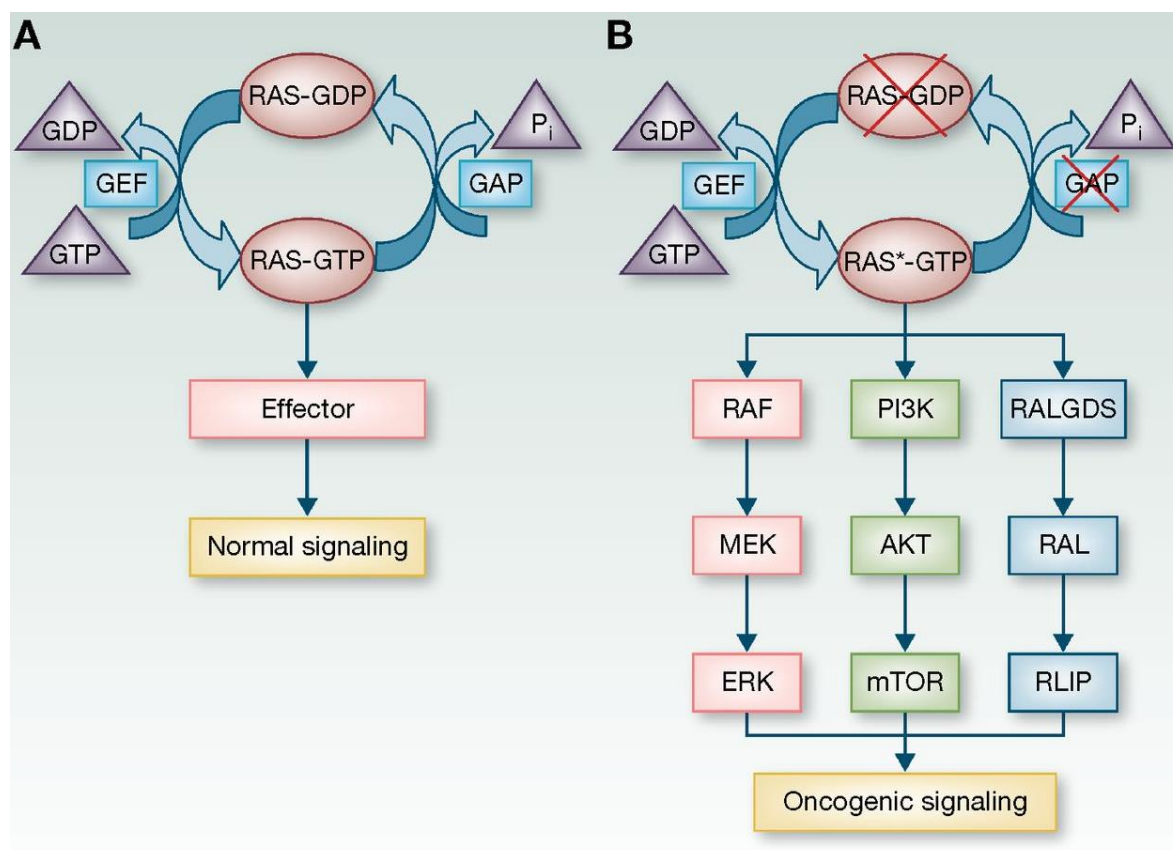


**Figure 3.6**: A, normal Ras signaling. B, oncogenic Ras signaling. When Ras is mutated, it is constitutively bound to GTP such that its GAP can not bind. The activated Ras signals through a multitude of effectors and downstream signaling pathways, a subset of which is shown here[120].

In addition, one group managed to develop irreversible inhibitors targeting K-Ras G12C, which is a common mutant found in K-Ras where the glycine at the twelfth position is mutated to cysteine. With submicromolar affinity, these inhibitors were shown to covalently bind K-Ras G12C and block GEFs binding, thus favoring the binding of GDP instead of GTP, and locking the K-Ras protein in its inactive state[121, 122]. More important, they exhibit antiproliferative effects and activate apoptosis in a K-Ras G12C–specific lung cancer cell line. Obviously, although these approaches are promising, more potent drugs that bind with even

higher affinity, say within nanomolar range, would be required for a clinically useful drug. In spite of that, these studies provide novel lead compounds for future optimization.

# Chapter 4. Methodology

CADD is an effective strategy for expediting and cost saving the drug discovery and development process. The significant gain in knowledge and structure information of both biological macromolecules and small molecules facilitate CADD to be extended and broadly applied to almost every stage in the drug discovery and development stage, from target identification and validation, to lead discovery and optimization, and preclinical tests.

In this thesis, we have used different CADD approaches to investigate the detailed structural mechanisms of the target proteins and search for inhibitors. There are many approaches in the field of CADD, the theoretical background of the methods applied in my Ph.D projects will be described in this chapter.

## 4.1 Homology model

It is necessary to have three-dimensional (3D) protein structures for the rational design of various types of biological experiments, such as search for specific inhibitors based on SBDD or site-directed mutagenesis. However, the number of available protein structures is not very large (~ 110000 in the protein data bank) in comparison to the number of known protein sequences, especially in case of transmembrane proteins. Therefore, when there is no available 3D structure for the target protein, various computational methods for generating 3D structures of proteins have been developed to resolve this issue. Based on a template with a similar sequence, homology modelling is able to generate the target structure. The idea behind this is that the number of possible folds in nature appears to be limited [123], thereby, the 3D structure of proteins is much more conserved than their sequences. For a given protein sequence, it is often possible to identify a homologous protein with a known structure. Homology modelling has proven to be reliable to generate a 3D model of a protein from its amino acid sequence if there is available template structure sharing more than 25% sequence similarities[124]. Building a homology model mainly consist four steps: the first step is to identify and select qualified structural template(s); the second step is to align target and template sequences; the third step is to predict the secondary structure and build model based on the template structure and the final step is to refine the model and do quality evaluation. The above steps can be repeated until a satisfying 3D model of target structure is achieved (Figure 3.1).
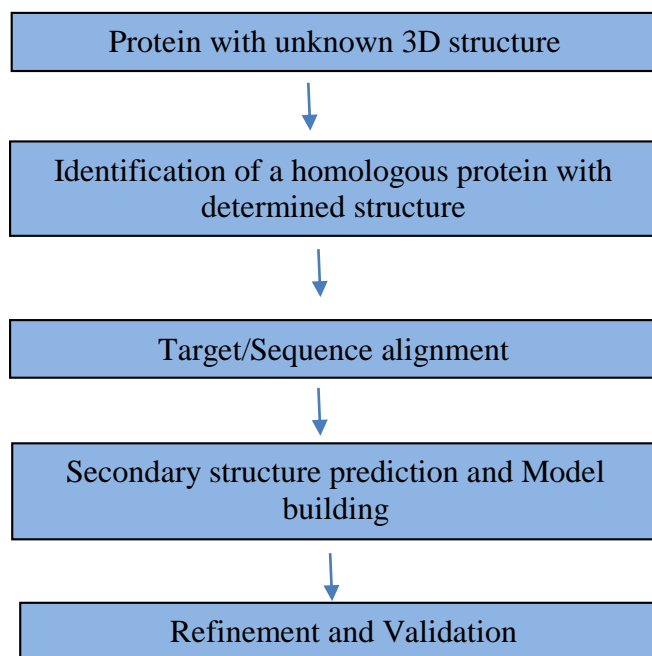
```
┌─────────────────────────────────────────────────┐
│          Protein with unknown 3D structure        │
└─────────────────────────────────────────────────┘
                        │
                        ▼
┌─────────────────────────────────────────────────┐
│      Identification of a homologous protein with  │
│              determined structure                 │
└─────────────────────────────────────────────────┘
                        │
                        ▼
┌─────────────────────────────────────────────────┐
│              Target/Sequence alignment            │
└─────────────────────────────────────────────────┘
                        │
                        ▼
┌─────────────────────────────────────────────────┐
│      Secondary structure prediction and Model     │
│                     building                      │
└─────────────────────────────────────────────────┘
                        │
                        ▼
┌─────────────────────────────────────────────────┐
│              Refinement and Validation            │
└─────────────────────────────────────────────────┘
```

**Figure 4.1:** Illustration of the homology modelling process.

Generally speaking, the accuracy of the homology model depends on the sequence similarity between target protein and templates, as well as the quality of the template structures. A template protein is usually found through sequence comparison with proteins in the Protein Database Bank (PDB) using algorithms such as BLAST (Basic Local Alignment Search Tool)[125] or FASTA (FAST-AII)[126] followed by alignment corrections. The structure of the target protein is then built by first copying the coordinates of the template structure that was used in the alignment. For residues that are different between the two proteins, only the backbone coordinates are copied, the side-chain coordinates are copied directly from the template as well if the aligned residues are the same and the sequence identity is high. Side-chains that are not identical to the template in the alignment are constructed using rotamer libraries and scored using energy functions. The next step, in most cases, is loop modelling, which is necessary if there are regions of the target sequence that are not aligned to or missed from the template. This step is most likely to generate modelling errors, especially if the loop is longer than 10 residues. The final generated homology model is optimized or energy minimized by force field methods, and validated for different purposes[127].

There are a number of software tools available for homology modelling, of which the mostly widely used are prime[128], SWISS-MODEL[129], MOE[130], MODELLER[131], and ROSETTA[132], etc. All those programmes are able to produce reasonable models when sequence identities

are above 30%[133]. In my studies, MOE and prime were mostly applied to generate homology models.

## 4.2 Docking

Molecular docking is a widely used method to predict the binding orientation of a molecule with respect to its specific target, typically protein or DNA in drug design. Mostly docking is performed either to study how a specific ligand interacts with a protein or to search a database of compounds for potential agents that can bind to a target protein. The increasing number of protein structures available in the PDB enables a large number of proteins to be considered as targets for therapeutic purpose. Docking can be divided into two main steps: initial posing of a ligand in an active site with the application of docking algorithms, followed by application of scoring function to assess the strength of the binding pose.

There are a large number of docking algorithms for posing a ligand in an active site available. Docking algorithms in early days did not treat the ligand and protein as flexible objects, hence only the six translational and rotational degrees of freedom were included[134]. Nowadays, we apply more reliable methods which involve flexible docking that the protein is treated as fixed during the docking; however, the ligand is able to move in order to take into account of the ligand's conformational degrees of freedom. In this case, the active site is not considered to undergo any significant conformational changes upon binding of a ligand. This type of docking is widely used with parallel computing resources to quickly and relatively accurately search databases for potential ligands to a target protein. The more accurate algorithms that consider flexibility of both the receptor and the ligand are extremely time consuming, thereby, have not been extensively developed[135]. The algorithms that treat ligand flexibility can be divided into three basic categories, including systematic methods, random or stochastic methods and simulation methods[135].

The docked poses are ranked and evaluated using scoring functions that approximate the binding free energy of a ligand to a receptor, which is a crucial step to differentiate correct poses from incorrect ones. The scoring functions make various assumptions and simplification in the evaluation of binding free energy for modelled complexes and do not fully consider some of the physical phenomena that are important for molecular recognition, i.e, entropic effects, as the scoring functions must be calculated rapidly during the docking run. Normally the scoring functions are expressed as a sum of separate terms that describe the various contributions to ligand binding. A large number of scoring functions are available,

such as forcefiled-based, empirical and knowledge-based scoring functions, which differ in which terms that are included in the expression of the binding free energy. Terms expressing nonbonded interactions, including van der waals interactions and electrostatic interactions, and solvation effects are commonly included [135].

## 4.3 Molecular dynamic simulation

Molecular dynamics simulation is an N-body simulation method for studying the physical movements of atoms and molecules, based on Newton's equations of motion to generate trajectories. This technique can be applied to material science, as well as biophysics and biochemistry to refine three-dimensional structures of proteins and other macromolecules. According to Newton's equations of motion F = m$a$, which states that a body with mass m, on which a force F is acting, experiences acceleration $a$ in the same direction as the force. The trajectory of an MD simulation is obtained by solving the differential equations in Newton's equations of motion. In each step of the simulation, a potential function V(r) is used to evaluate the total force acting on each particle and is obtained from Equation 4.1, which is calculated using a molecular mechanics forcefield.

$$V(r) = \sum_{bonds} \frac{k_b}{2} \left(l_i - l_{i,0}\right)^2 + \sum_{angles} \frac{k^\theta}{2} \left(\theta_i - \theta_{i,0}\right)^2 + \sum_{torsions} \frac{V_n}{2} [1 + \cos(n\omega - \gamma)] +$$
$$\sum_{i=1}^{N} \sum_{j=i+1}^{N} 4\varepsilon_{ij} \left(\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right) \qquad 4.1$$

As stated from the above equation, bond stretching, angle bending, bond rotation (torsion), and non-bonded interactions, including van der Waals interactions and electrostatic interactions, contribute to the total V(r). Bond stretching is expressed by a harmonic function about the reference bond length $l_{i,0}$, which represents the bond length when all other terms in the force field are set to zero. However, a harmonic Morse potential could be more appropriate to apply when the system is far away from equilibrium. Angle bending is also expressed by a harmonic function about the reference bond angle $\theta_{i,0}$, and the expression of angle bending is defined for all combinations of three bonded atoms. The terms $k_b$ and $k_\theta$ are force constants for bond stretch and angle bend, respectively. The torsional function describes the dihedral angle between four bonded atoms and is calculated using cosine functions. In the torsion expression above, $V_n$ is the barrier height, n is the number of minima in the function when the bond is rotated 360°, $\omega$ is the torsion angle, and $\gamma$ is the phase factor that determines

where the minima are located. The last expression in the above equation is non-bonded interaction, it sums the van der Waals interactions and electrostatic interactions of all the particle pairs. van der Waals interactions is commonly described using the Lennard Jones 12-6 potential, which contains the following parameters; $\varepsilon$ is the depth of the potential well, $\sigma$ is the finite inter-particle distance at which the potential is zero, and r is the distance between particles. The electrostatic interactions are usually calculated as a sum of interactions between pairs of point charges q using Coulomb's law. In periodic systems, it is common to sum the electrostatic interaction by using the Ewald summation method, which splits in two terms, one in real space and one in reciprocal space, which results in faster convergence. An extension of the Ewald summation is the particle mesh Ewald (PME) summation[136, 137].

The total force F is then calculated as a negative gradient of the potential energy function, and further used to calculate the acceleration of the particles according to Equation 4.2-4.3,

$$F_{r_i} = -\frac{dV_r}{dr_i} \qquad 4.2$$

$$\frac{d^2 r_i}{dt^2} = \frac{F_{r_i}}{m_i} \qquad 4.3$$

Together with the current positions and velocities, the acceleration is used to calculate the positions and velocities in the next step of the simulation. The initial position can be obtained from the simulated structures, and the initial velocities can be generated according to atom types and simulation temperature.

Several algorithms for integration of the equations of motion exist, all which apply Taylor series expansions as approximations of the positions, velocities and accelerations. The two most widely used, the leapfrog algorithm[138] and the Verlet algorithm[139] are expressed in Equation 4.4 and 4.5 respectively,

$$R(t + \Delta t) = R(t) + \Delta t \cdot v\left(t + \frac{1}{2}\Delta t\right)$$

$$4.4$$

$$v\left(t + \frac{1}{2}\Delta t\right) = v\left(t - \frac{1}{2}\Delta t\right) + \Delta t \cdot a(t)$$

$$R(t + \Delta t) = R(t) + \Delta t \cdot v(t) + \frac{1}{2}\Delta t^2 \cdot a(t)$$

$$v(t + \Delta t) = v(t) + \frac{1}{2}\Delta t \cdot [a(t) + a(t + \Delta t)]$$

4.5

The length of the time step, $\Delta t$, used in a simulation is required to be significantly less than the period of the fastest motion in the system, such as vibrations of bonds. The time step is normally set to be 1-2 femtoseconds.

In order to better mimic experimental conditions of the system in MD simulations, keeping the temperature and pressure constant is important. Different statistical ensembles are common setups in MD simulations to control these physical properties. These statistical ensembles can be generated based on which state variables, such as the energy E, volume V, temperature T, pressure P, and number of particles N, are kept fixed. NVE ensemble, also referred to as the microcanonical ensemble, keeps the energy E and volume V constant. Thus there is no temperature and pressure control in this ensemble, which is not recommended for equilibration. NVT ensemble, also known as canonical ensemble, keeps both temperature and volume constant throughout the run, in which the temperature is kept constant through direct temperature scaling and temperature bath coupling. NPT ensemble allows both the temperature and pressure constant, where the pressure is adjusted by volume adjustment. The number of particles is conserved in all ensembles. Either NVT or NPT was carried out in the simulations of the work in this thesis.

To be concluded, the whole MD workflow can be simplified as in Figure 4.2: in a classical MD simulation, initial positions $R_0$ and velocities $v_0$ as well as a time step $\Delta t$ (normally 1-2 fs) must be provided. $R_0$ can be obtained from X-ray or NMR structures, and $v_0$ value is usually generated according to atom types and simulation temperature. A potential function V(r) is further used to evaluate the force on each atom, in order to get accelerations $a$. New position $R_{n+1}$ can then be obtained from $R_n$ repeatedly. When a sufficiently long trajectory is generated, various properties of the system can be analysed, such as root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), hydrogen bonds, radius of gyration, distances, solvent accessible surface areas, vibrational motions, etc, which provides useful information about the protein-ligand interactions, protein structural fluctuations and so forth.
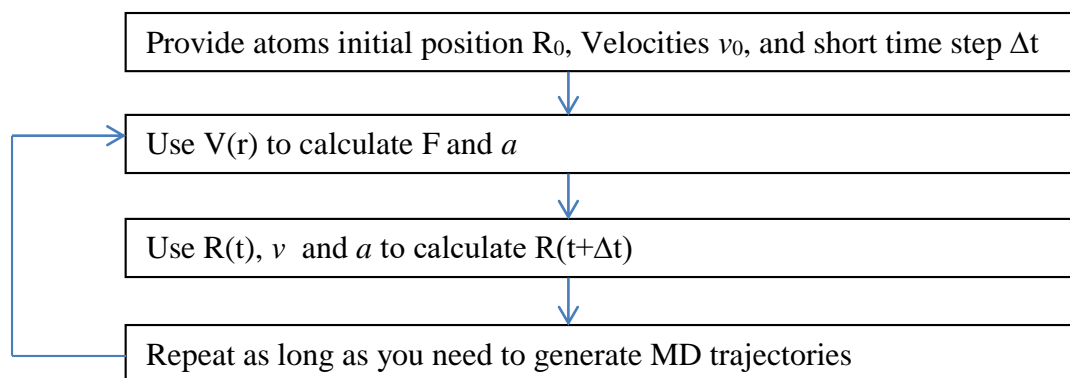
| Provide atoms initial position $R_0$, Velocities $v_0$, and short time step $\Delta t$ |

| Use V(r) to calculate F and $a$ |

| Use R(t), $v$ and $a$ to calculate R(t+$\Delta$t) |

| Repeat as long as you need to generate MD trajectories |

**Figure 4.2:** MD steps

## 4.4 MM-PB(GB)SA approach

The Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) and the Molecular Mechanics Generalized Born Surface Area (MM-GASA) methods are commonly used to evaluate the binding free energies between two molecules from MD simulations, normally sets of complex structures are collected from MD trajectories for the calculation. The binding free energies calculated by MM-PB(GB)SA methods are evaluated according to the Equations 4.6-4.7.

$$\Delta\Delta G_{bind} = \Delta G_{complext} - (\Delta G_{protein} + \Delta G_{lignd}) \qquad 4.6$$

Each term can be estimated in the following:

$$\Delta G = \Delta G_{MM} + \Delta G_{sol} - T\Delta S$$
$$\Delta G_{MM} = \Delta G_{ele} + \Delta G_{vdw} \qquad 4.7$$
$$\Delta G_{sol} = \Delta G_{PB/GB} + \Delta G_{SA}$$

Where $\Delta G_{MM}$ represents the molecular mechanics free energy, which includes the electrostatic interactions $\Delta G_{ele}$ and van der Waals interactions $\Delta G_{vdw}$. $\Delta G_{sol}$ is the solvation free energy that consists of polar contributions of electrostatic solvation energy $\Delta G_{PB/GB}$ , and non-polar contributions of the non-electrostatic solvation, $\Delta G_{SA}$. The conformational change upon ligand binding, $T\Delta S$, can be estimated with the normal mode analysis on a set of complex structures obtained from MD simulations.

## 4.5 Structure based-pharmacophore

Screening a 3D database against a pharmacophore hypothesis is generally more computationally efficient than structure-based docking, which involves many energy evaluations as part of the conformational searching and scoring process. Recently, methods

have emerged that attempt to take advantage of both the speed of pharmacophore screening and information of structure-based docking. In doing so, methods have been developed to generate pharmacophore hypotheses derived from protein-ligand complexes[140-142]. Study has shown to discover novel leads for 11β-hydroxysteroid dehydrogenase type 1 (11β-HSD1) enzyme based on these methods[143]. In this thesis, I used a novel protocol for generating energy-optimized pharmacophores (e-pharmacophores) based on mapping of the energetic terms from the Glide XP scoring function onto atom centers of the ligand located in the binding pocket. The Glide XP scoring function is presented in the following Equation 4.8:

$$XP\ GlideScore = E_{coul} + E_{vdW} + E_{bind} + E_{penalty}$$

$$E_{bind} = E_{hyd\_enclosure} + E_{hb\_nn\_motif} + E_{hb\_cc\_motif} + E_{PI} + E_{hb\_pair} + E_{phoic\_pair} \qquad 4.8$$

$$E_{penalty} = E_{desolv} + E_{ligand\_strain}$$

The advantages of this scoring function is that it includes more complex energy terms than traditional molecular mechanics or empirical scoring functions, such as hydrophobic enclosure ($E_{hyd\_enclosure}$), special neutral−neutral hydrogen-bond motifs ($E_{hb\_nn\_motif}$), $\pi$ stacking and $\pi$-cation interactions ($E_{PI}$), standard ChemScore-like hydrogen bond ($E_{hb\_pair}$) lipophilic pair ($E_{phoic\_pair}$) rapid docking of explicit waters ($E_{desolv}$), contact penalties ($E_{lig\_strain}$).These terms are described in detail in the original Glide XP work[144].



```
┌─────────────────────────────────────────────┐
│            Structure preparation            │
└─────────────────────────────────────────────┘
                      │
                      ▼
┌─────────────────────────────────────────────┐
│              Ligand refinement              │
└─────────────────────────────────────────────┘
                      │
                      ▼
┌─────────────────────────────────────────────┐
│   Map GlideXP descriptor on to pharmacophore site   │
└─────────────────────────────────────────────┘
                      │
                      ▼
┌─────────────────────────────────────────────┐
│             Hypotheis generation            │
└─────────────────────────────────────────────┘
                      │
                      ▼
┌─────────────────────────────────────────────┐
│            Find mathes to hypothesis        │
└─────────────────────────────────────────────┘
```
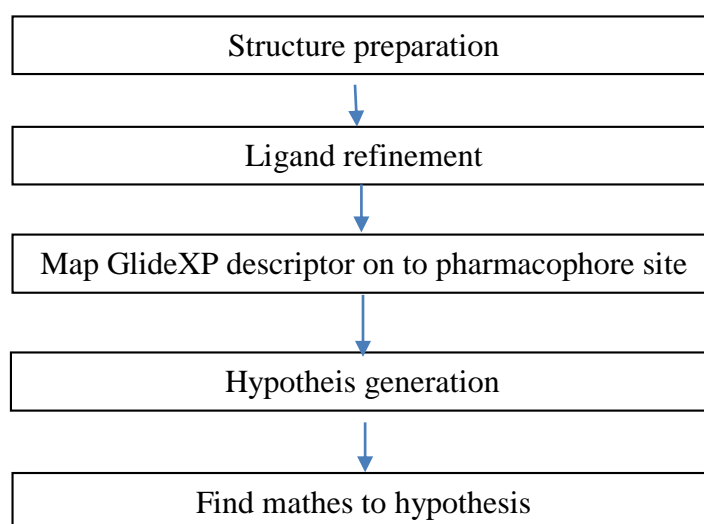
**Figure 4.3:** Structure based pharmacophore steps

The whole workflow for generating a structure-based pharmacophore begins with a ligand−receptor complex, refinement of the ligand pose, computing the Glide XP scoring terms, and mapping the energies onto atoms. Then, pharmacophore sites are generated, and

the Glide XP energies from the atoms that comprise each pharmacophore site are summed. The sites are then ranked based on these energies, and the most favorable sites are selected for the pharmacophore hypothesis. Finally, these e-pharmacophores are used as queries for virtual screening (Figure 4.3)[145].

## 4.6 Ligand-based pharmacophore and 3D-QSAR modelling

Pharmacophore modelling has been extensively used in drug discovery, due to its ability to both find and optimize active molecules[146-148]. Pharmacophore modelling is based on the general concept that molecules share similar arrangements of related chemical groups, such as hydrogen-bond donors or acceptors, or aromatic rings, in spatially and geometrically can make comparable interactions with a receptor, hence those compounds have similar chemical and biological activity. It may be possible to generate a common pharmacophore model when chemical knowledge of many active ligands is available. This model in principle represents the key molecular features necessary for activity, including lipophilic, aromatic, hydrogen bonding, and charged groups. Once the pharmacophore model is generated, it can then be used to screen compound databases for molecules that share the same features. Pharmacophore modelling is most commonly carried out when the target structure is not available. Some targets in drug discovery campaigns are very difficult to crystallize, such as ion channels, transporters, or G protein-coupled receptors (GPCRs), it is thus particular useful to generate pharmacophore models to search for potential hits[149]. Further, if the activity data of the ligands, such as $IC_{50}$ values, are also known, then it can be applied to optimize the hypothesis and develop a 3D-QSAR model for further predicting the activity of the newly found compounds.

In the following, I will briefly introduce how to build a pharmacophore and develop 3D-QSAR model in the PHASE module[150]: The first step involves preparing the ligands, the ligands begin with a structure cleaning step by LigPrep, followed byconformational space search by torsion sampling or mixed MCMM/LMOD[151, 152] method. The minimized structures that are ultimately obtained are then filtered through a user-defined relative energy window. The second step deals with creating pharmacophore sites: Six built-in types of pharmacophore features are available: hydrogen bond acceptor (A), hydrogen bond donor (D), hydrophobe (H), negative ionizable (N), positive ionizable (P), and aromatic ring (R). In addition, users may define up to three custom feature types (X, Y, Z) to account

for characteristics that don't exist in the six built-in types. The third step is perceiving common pharmacophores, which can be done based on using a tree-based partitioning
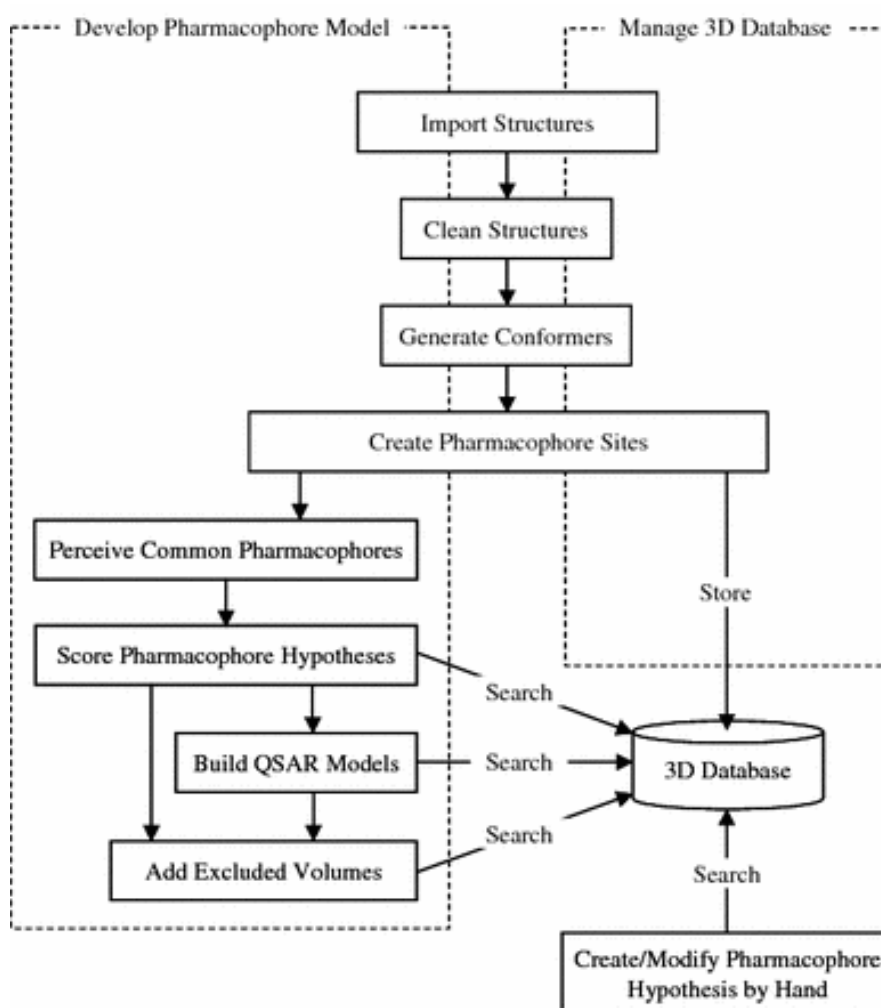


**Figure 4.4:** Workflow of generation of ligand-based pharmacophore and 3D-QSAR model[150]

technique that puts similar pharmacophores into a group according to their intersite distances. As the fourth step pharmacophore hypothesis is scored. The pharmacophore model can be scored with respect to actives, and thus filter out inappropriate pharmacophores and elevate pharmacophore models with essential features for high binding affinity. The fifth step is about building 3D-QSAR model. In order to develop 3D-QSAR model for each hypothesis, it requires sufficient number of molecules of various activity that can be used as training set structures and employed fitness. A QSAR model can be either atom-based or pharmacophore-based, in atom-based QSAR, all atoms of a

molecule are taken into account, while in pharmacophore-based QSAR, only the pharmacophore sites are considered. The QSAR models are generated by applying partial least squares (PLS) regression, with the maximum PLS factors no larger than 1/5 the number of training set molecules.

## 4.7 Density functional theory

Density functional theory (DFT) is presently the most successful quantum mechanical modeling method used in physics and chemistry to compute the electronic structure (principally the ground state) of many-body systems, in particular atoms, molecules, and the condensed phases. In chemistry, DFT is used to predict a variety of molecular properties, such as molecular structures, vibrational frequencies, atomization and ionization energies, electric and magnetic properties, reaction paths, etc. The modern DFT calculations are based on two Hohenberg and Kohn theorems, which proves that the electronic energy of a molecule in a ground state could be determined completely by electron density $\rho(\mathbf{r})$[153]. The electron density $\rho(\mathbf{r})$ can be defined as in Equation 4.9, where r is spatial variable of electrons and s is the spin variable of electrons.

$$\rho(r) = N \sum_{s_1} \dots \sum_{s_N} \int dr_2 \dots \int r_N |\Psi(r_1, s_1, r_2, s \dots r_N, s_N)|^2$$

$$\int \rho(r)dr = N$$

4.9

The Kohn-Sham (KS) theories are the most common implementation of DFT, making it widely used. The KS equations are analogous to the Hartree-Fock equations. In the KS model, non-interacting electrons moving in an effective potential is introduced to solve the problem of interacting electrons of many-body moving in a static external potential. The most popular DFT method is the Becke3-Lee-Yand-Parr(B3LYP) hybrid functional[153], and was also used for the calculations in this thesis. Generally speaking, DFT is not a CADD method, however, it is involved in application in CADD to predict molecular properties.

# Chapter 5. Summary of papers

In collaboration with different experimental groups, we have investigated various protein targets, which all have in common that they overexpressed in cancer cells. Therefore, CADD approaches were applied to either investigate the structural conformations or develop inhibitors to the targets. The targets studied includes RET, Eg5, KIF18B, Tip60 and K-Ras.

## 5.1 Studying DFG-out inhibitors targeting RET (Paper I)

In paper I, a range of different CADD approaches were employed. Homology modelling was used to predict the DFG-out conformation of RET tyrosine kinase domain, followed by docking to predict the binding mode of DFG-out inhibitors to the RET. Based on the complex structures, MD simulations were carried out to further optimize the structures and calculate the MM-PBSA interaction energies. Finally, the key features of a structure-based pharmacophore were determined based on the complex structures.

RET is a transmembrane tyrosine kinase receptor, and overexpression of RET are related to several different thyroid cancers. In this study, the binding mode of four DFG-out tyrosine kinase inhibitors with different IC50 values, namely Abt-348, Birb-796, Motesanib and Sorafenib, to a homology model of the DFG-out conformation of RET were explored (Figure 5.1). During MD simulations, Abt-348, Birb-796, Motesanib, Sorafenib, all form stable hydrogen bonds with Glu775 and Asp892 of the ATP binding pocket of RET. Furthermore. All four ligands form stable hydrogen bonds in the hinge region, Abt-348 hydrogen bonds to the Glu805 and Ala807 backbone, the other three inhibitors formed stable hydrogen bonds to Ala807. The hydrogen bonds highlight that the important interactions between RET and the inhibitors. The MM/PB(GB)SA analyses of Abt-348, Birb-796, Motesanib, and Sorafenib interaction with RET were further conducted to highlight their detailed structural interactions. Through the analysis of the components in binding free energy, it suggests that the van der Waals energy and the nonpolar solvation energies, which are responsible for the burial of the inhibitors hydrophobic groups in the cavity, are the major contribution to the binding. However, the electrostatic contributions in vacuum and solvent ($\Delta G_{ele} + \Delta G_{ele,sol}$) constitute the unfavourable part to binding. The further residue free energy decomposition calculations showed that the major favorable energy contributions originate predominantly from residues Glu775, Ile788, Val804, Tyr806, Ala807, Leu881, Ile890, Ser891 and Asp892. In addition, we generated e-pharmacophore models of the RET-inhibitor complexes, which all shared the common

features D–A–R–A, where R is aromatic ring, A is hydrogen acceptor and D is hydrogen donor. The RET-Abt-348 complex structure showed ten features, of which three were aromatic rings, three hydrogen acceptors and four hydrogen donors. Six features were identified for the RET-Birb-796 complex structure, including two ring aromatics, two hydrogen acceptors and two hydrogen donors. Another six features, of which three are ring aromatics, two are hydrogen acceptors and one is hydrogen donor, were obtained for the RET-Motesanib complex. Finally, eight features that contain three ring aromatics, two hydrogen acceptors and three hydrogen donors were shown for the RET-Sorafenib-I complex structure. Taken together, the results herein enlightened the protein–ligand interactions between RET and tyrosine kinase DFG-out inhibitors, which can be used as a basis for future rational design of novel potent inhibitors to RET.
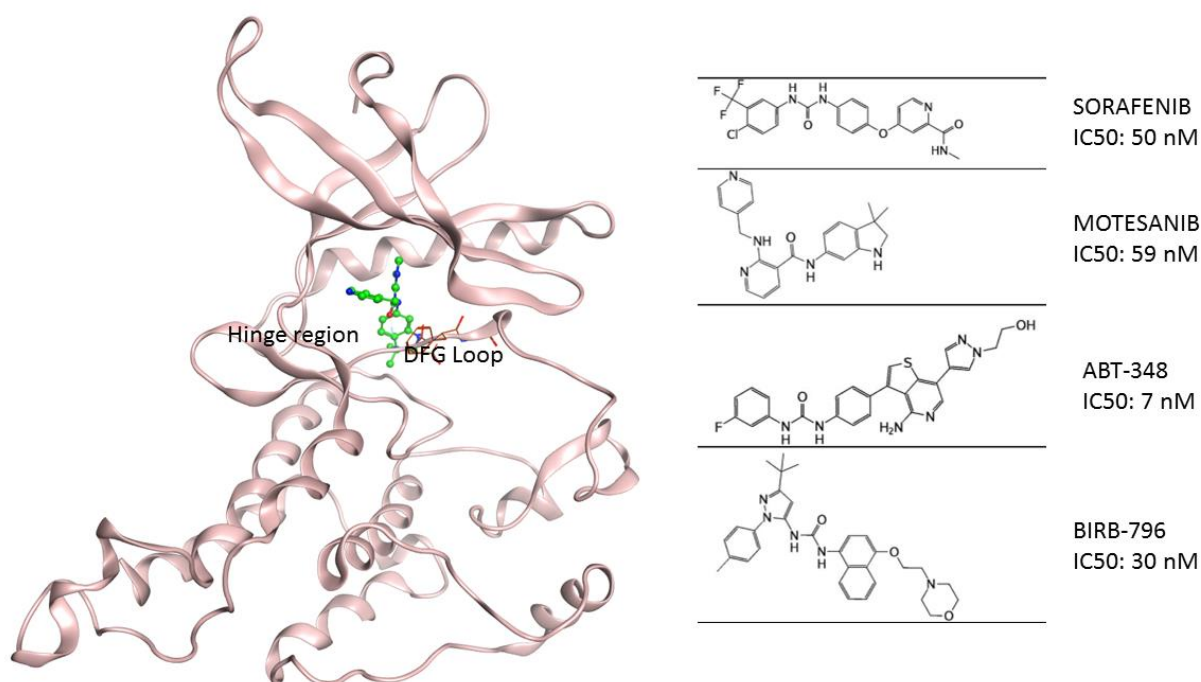


**Figure 5.1:** RET with Motesanib binding in its ATP binding pocket. Hinge region and DFG loop is labelled. Four inhibitors with different $IC_{50}$ values are shown.

## 5.2 Investigating cadherin like domain of RET. (Paper II)

As it is well known that the calcium binding site could be an appropriate structural target in tumours with wildtype RET expression, the aim of paper II is to investigate the structural details of how calcium depletion or specific mutations around the calcium binding site, namely R231H, D264K, and D300K, influence the RET CLD1-4

48

conformation, and to present the design of inhibitors targeting this region. RET CLD was firstly constructed by homology modelling, followed by *in silico* mutagenesis, and 300 ns molecular dynamics (MD) simulations of the RET CLD1-CLD4 wildtype structure, the calcium-depleted structure, and structures carrying with mutations commonly found in Hirschsprung's disease (HSCR), namely R231H, D264K, and D300K (Figure 5.2)[4,11]. We furthermore present a structure-based pharmacophore based on the most populated structure obtained from a cluster analysis of the MD simulations of the wildtype systems, to provide a first tool for designing inhibitors of the RET extracellular domain.
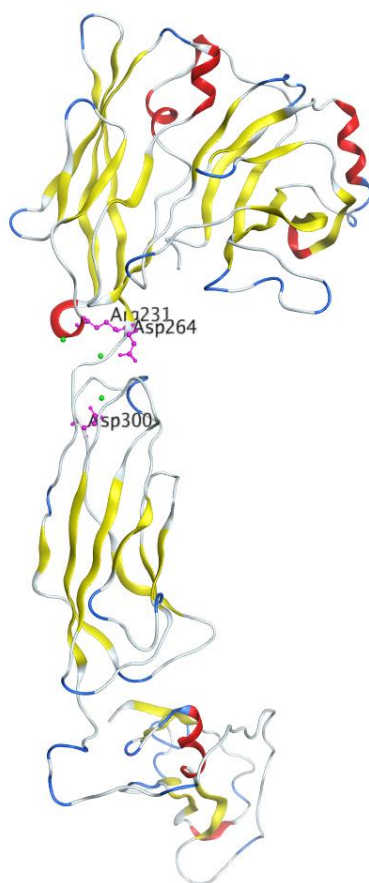


**Figure 5.2:** RET CLD structure, with mutant residues labelled in magenta and calciums in green spheres.

The structural details of the conformational changes of RET CLD1-4 upon calcium-depletion and mutations in the calcium-binding region were explored, using all-atom MD simulations to quantitatively analyse properties and explain the role of the mutants in HSCR pathology. Upon mutation or depletion of calcium in RET CLD1-4, several common features were noted,

such as large fluctuations in the RMSF, a decrease in radius of gyration, and an increase in distance between calcium ions. In addition, PCA revealed that the dynamic motion covers a larger region of phase space along the PC2 axis for wildtype, but a larger region of phase space along PC1 for the non-calcium-binding and the mutant RET CLD1-4. The PC1 corresponds to motions of the CLD1-2 swinging out, CLD3 moving inwards, and CLD4 elevating up; PC2 corresponds to the motion of the CLD1-2 swinging out, CLD3-4 extending down. The R231H and D264K mutant systems display more similarities to wildtype and result in smaller structural changes, whereas the calcium-depleted system and the D300K mutant demonstrate significant structural changes. The mechanism leading to protein degradation and non-ligand interaction hence most likely differs for these systems. We propose that the small structural changes in R231H and D264K mutants might propagate changes through allosteric mechanisms that finally destroys the protein conformation, whereas the significant structural changes in the calcium depleted or D300K mutant system result in destroyed protein conformation, leading to the malfunction of the RET CLD1-4. Furthermore, based on fragment docking and clustering we propose a pharmacophore for possible wildtype RET CLD1-4 inhibitors targeting the extracellular ligand contact site. Based on this pharmacophore we further screened an in house compound database, of which two compounds showed effective inhibitory targeting RET. However, the structures of the compounds will not be shown here for the potential future patenting.

## 5.3 Understanding a photo-switchable inhibitor targeting RET (Paper III)

A photoswitchable RET kinase inhibitor was developed based on azo-functionalized pyrazolopyrimidines to gain external control of the activity of RET (Figure 5.3). It displays excellent switching properties and stability with good inhibitory effect towards RET in cell-free as well as live-cell assays and a moderate difference in inhibitory activity between its two photoisomeric forms. To further investigate its photoswitchable properties in the RET binding pocket, docking, time dependent density functional theory (TDDFT) calculations, and MD simulations were applied to explain how the isomerization affects the structural changes in the RET tyrosine kinase that lead to the ligand affinity differences.

In the TDDFT calculations, the absorption wavelength maxima of the inhibitor in its E isomer is 330 nm, and 296 nm in its Z isomer, close to the experimental value 349 nm and 299 nm respectively. The E and Z isomer of the inhibitor was docked into the active site of RET, and was noticed that the Z isomer could barely fit into the pocket. However, the difference in binding affinity between E and Z isomers is only 3-4 folds, in favour of E4

binding. Therefore, the docking pose of the Z isomer should not be trusted. We thus carried out MD simulations of the E isomer binding to the pocket, and applied a torsional force to twist the dihedral C-N=N-C into the Z isomer, thereby achieving a proper binding pose of Z-isomer. This was followed by two parallel MD simulations of 500 ns were performed on the RET-E isomer and RET-Z isomer complexes. After 15 ns, a stable Z isomer in the pocket was observed. Approximate interaction energy between RET and the E/Z isomers was calculated, which showed a 30 kJ/mol difference. The conclusion of this paper is that the TDDFT calculations can be further applied to predict the optical properties of similar photoswitchable compounds; the Z-4 isomer interferes the inhibitor-receptor interactions rather than unable to enter the active site; the favoured Z-4 binding mode generated from MD simulations in the active site lead to the interaction energy increases and lowers the differences between E and Z isomer on binding to RET.
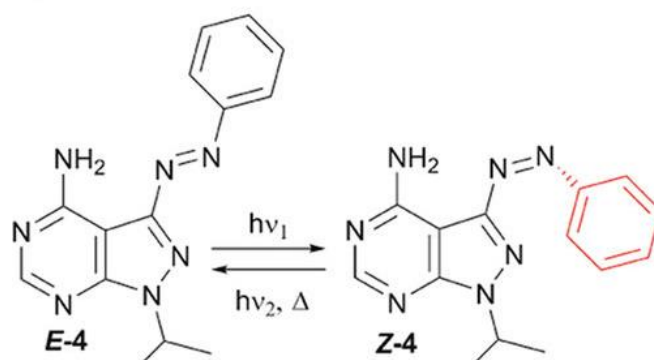


**Figure 5.3:** Structure and isomerization of photoswitchable compounds E4 and Z4[154].

## 5.4 Rational design a Tip60 inhibitor (Paper IV)

In paper IV, we aimed to develop inhibitors for Tip60 histone acetyltransferase, which is a potential therapeutic target in cancer treatment. This paper consists of theorectical calculations, synthesis, and experimental tests. We initially designed the inhibitors based on the pentamindine (PNT) scaffold, which has been reported to inhibit Tip60 activity by decreasing its H2A acetylation, and further optimized the structure using a combinatorial builder to enhance the interactions between the target and the designed inhibitors. We also carried out MD simulations to observe the stability of different inhibitors in the binding pocket.

Homology model of Tip60 indicated that one end of the binding pocket is mainly positive, which arises from side chains Arg326, Lys331 and Arg434, whereas the other end is negative,

which is from the side chain of Glu351 (Figure 5.4). We docked the PNT to the Tip60 active site and *in silico* generated series of derivatives within the pocket based on the possible surrounding interactions. A novel Tip60 inhibitor, TH1834, was achieved to fit this specific pocket. Following synthesis of the compound, TH1834 was demonstrated to show significant inhibition activity towards Tip60 *in vitro* and treating cells with TH1834 results in apoptosis and increased unrepaired DNA damage in breast cancer but not control cell lines following ionization radiation treatment. Furthermore, TH1834 demonstrated its specificity to RET by not affecting the activity of related family member MOF, as MOF dependent post-translational acetylation of histone H4 at lysine 16 (H4K16ac) was not affected by TH1834. It was further shown that by manipulating Tip60 activity through inhibition using TH1834, the effect of ionizing radiation against cancer cell lines (MCF7, PC-3 and DU-145) was increased, resulting in apoptosis. Our data also suggests that Tip60 dependent pathway is necessary for cancer cell line survival. In conclusion, the modelling, synthesis and validation of the small molecule Tip60 inhibitor TH1834 in this study indicates a first step towards developing additional specific, targeted histone acetyltransferase inhibitors, which may lead to further improvements in breast cancer treatment. Our findings demonstrate the potential therapeutic applications of Tip60 inhibitors, and the future work will be focused on using this model to produce compounds with increased potency.
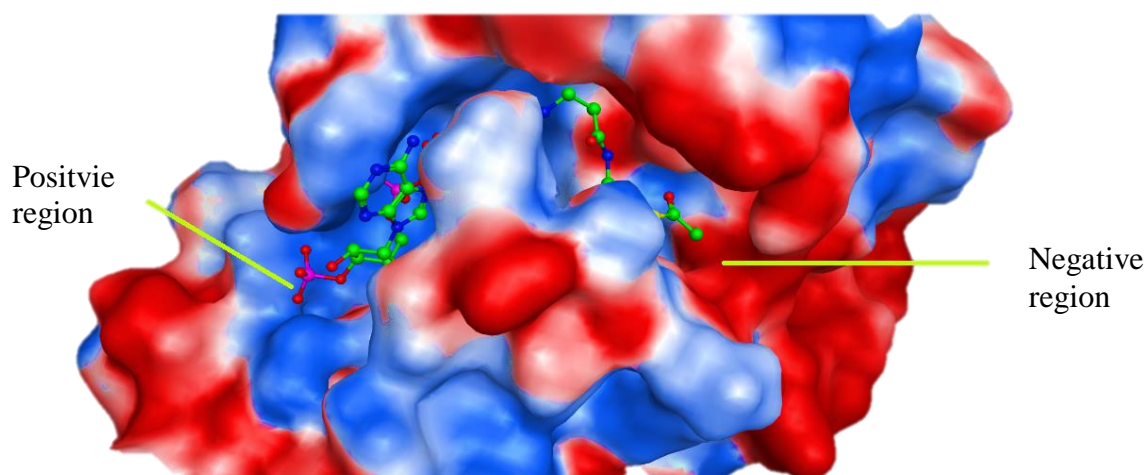


**Figure 5.4:** The electrostatic surface potential of the Tip60 binding pocket, the positive and negative region at each end of pocket is pointed out. Acetyl-CoA is located in the binding pocket.

**5.5 Analysis Biphenyl Type Inhibitors Targeting Eg5 (Paper V)**

The kinesin protein Eg5 (also known as Kinesin-5 or KIF11) plays an essential part in the formation and maintenance of the bipolar spindle. It is considered to be an attractive target for cancer chemotherapy with reduced side effects. In our study we performed a detailed computational modelling study on 66 biphenyl type inhibitors binding to the α4/α6 allosteric pocket of Eg5, as this type of inhibitors show a unique activity profile compared to the known α2/L5/α3 allosteric inhibitors. Due to crystal packing effects, loop L11, which is located in the entrance of the binding pocket, is missing in the available crystal structure. In order to gain insight into the role of this flexible loop, homology modelling followed by MD simulations were carried out to sample L11 conformations starting from different state. It was shown that biphenyl type inhibitor binding to Eg5 brings about less fluctuations of L11 and stabilizes its conformation. Analysing L11 conformation, we note that residue Asn287 from L11 forms hydrogen bonding to the sulfone group of co-crystalized PVZB1194 and moves inward to the pocket (Figure 5.5), whereas in the unbound protein structure L11 moves away from the α4/α6 region. Free energy analysis supports the effect on L11 by inhibitor binding.
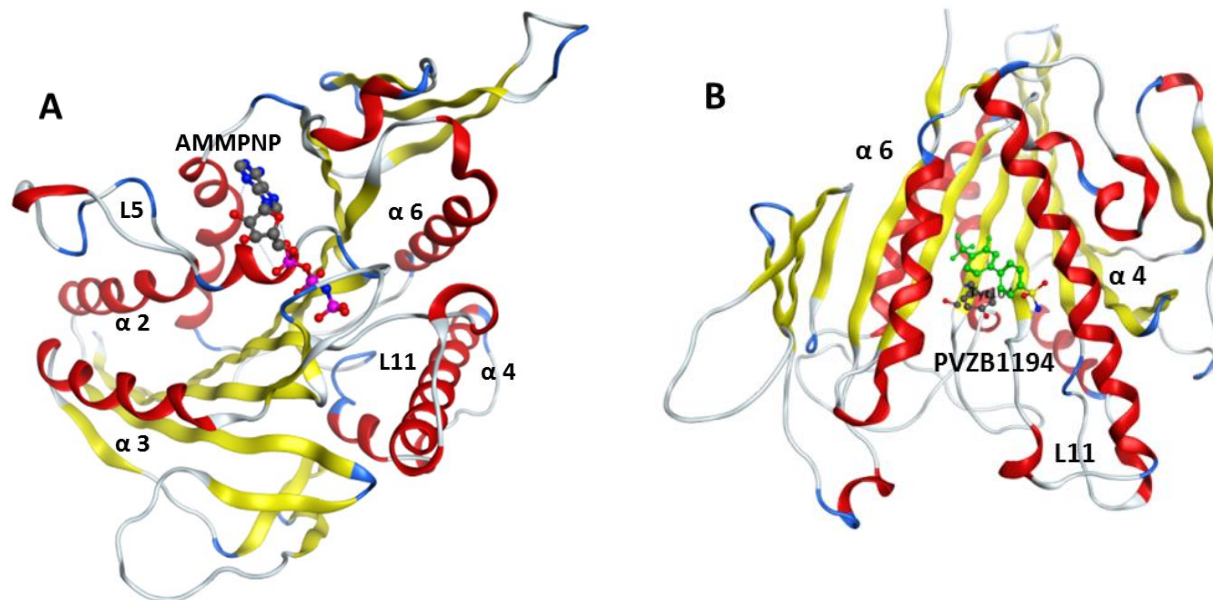


**Figure 5.5:** A  Eg5 with bound AMMPNP , showing pockets α2/L5/α3 and α4/L11/α6. B PVZB1194 bound to homology model Eg5 with ordered L11

We furthermore carried out pharmacophore, 3D-QSAR and pharmacokinetic modelling studies of a large set of biphenyl type inhibitors of Eg5. The best pharmacophore model

DDRRH.6 consists of two hydrogen bond donors (D), two aromatic rings (R) and one hydrophobic group (H). The pharmacophore DDRRH.6 has correlation coefficients, $R^2$=0.81 and $Q^2$=0.64 in 3D-QSAR model, indicating this model can further utilised to predict the activity of new compounds sharing the same pharmacophore features. The ligands were further studied with docking into the α4/α6 allosteric pocket using structure obtained from the MD simulations. In addition, a structure based pharmacophore generated from fragment docking to the allosteric pocket shares good overlap for the DHRR features. Extra hydrogen bond acceptors and hydrophobic groups identified in the structure based pharmacophore can provide further possibilities in order to explore the new regions of the active site. The study here provides a basis for further optimization or development of allosteric inhibitors targeting the α4/α6 pocket of Eg5.

## 5.6 Designing inhibitors targeting KIF18B (Paper VI)

KIF18B is one of kinesin 8 family members, which walk directionally towards the plus end of microtubules and modulate microtubules dynamics in spindle assembly and organization. It has been found that deregulation of KIF18B leads to carcinogenesis, although detailed understanding of the specific role of KIF18B is still lacking. The aim of this study is to design inhibitors targeting KIF18B to prevent its activity, to enable studies of its specific roles in normal cells vs cancer cells. By applying the molecular modeling approaches of homology modeling, docking, and pharmacophore hypothesis generation and screening, we showed that the inhibitors targeting Eg5, which is also a kinesin protein, do not fit into the binding pocket of KIF18B, neither the conventional L5/α2/α3 pocket nor α4/α6 pocket. We furthermore investigate the KIF18A inhibitor BTB-1 and its derivatives, since KIF18A and KIF18B share 60% of sequence similarity in motor domain, the KIF18A inhibitors turn out promising on targeting the KIF18B as well. Thus, based on a pharmacophore hypothesis generated from six BTB-1 derivatives, BTB-1, 3, 5, 6, 7, 13, which show the highest inhibitory activity, we screened 3.6 million in stock lead-like compounds, and 29000 compounds were achieved from this screening sharing the same pharmacophore key features to the active BTB-1 derivatives. In the next step, we performed docking of these obtained compounds to both KIF18A and KIF18B, and a final 10 compounds chosen for the further experimental validation. Fortunately, none of the final 10 binders were identical between KIF18A and KIF18B, thus providing promising selectivity of the proposed KIF18B inhibitors. The compounds showed promising inhibitory activity towards KIF18B, which is not included to be discussed here for the possible further patenting.

**5.7 Impact of mutations on K-Ras-p120GAP interaction (Paper VII)**

The K-Ras protein plays an important role in the signal transduction cascade of cell proliferation and/or differentiation. K-Ras binds guanosine triphosphate (GTP) in its active state and triggers the downstream signalling pathway. By using GTPase activating protein (GAP) to facilitate hydrolysing GTP to guanosine diphosphate (GDP), the K-Ras inactivation occurs. However, Certain mutations in K-Ras lead to a permanent active state with GTP binding which leads to tumorigenesis due to failed interaction with GAP. In this paper, we examined the mutations E31N, D33N and D38N located in K-Ras Switch I region, and mutation K935N of GAP-334 (Figure 5.6). When bound with K-Ras, K935 is spatially located to Switch I region in complex structure. We aimed to investigate the potential mechanism by which these mutants affect the interaction between the two proteins.
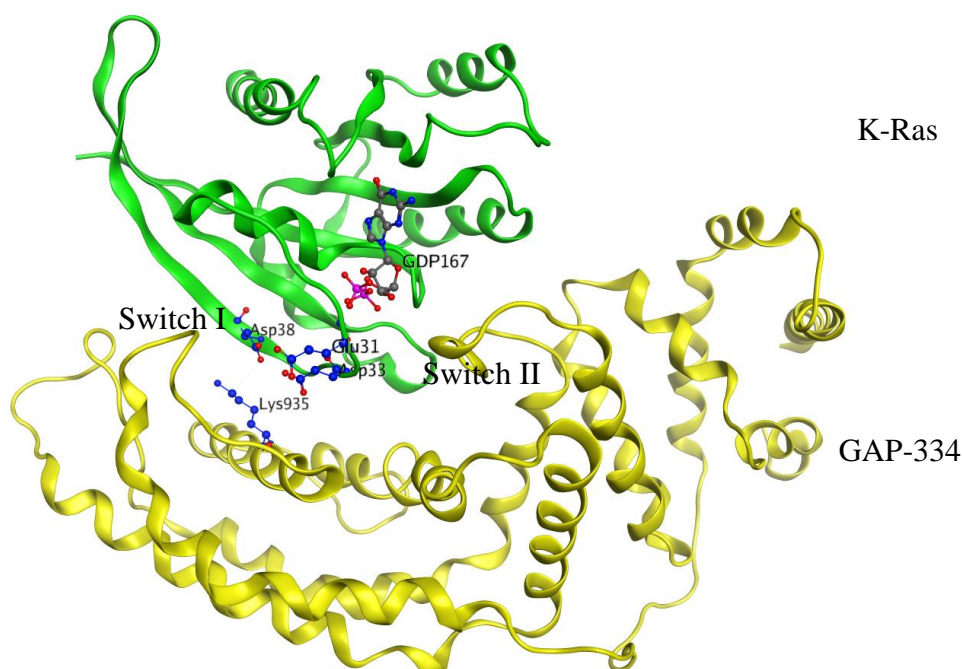


**Figure 5.6:** The complex between GAP-334 and K-Ras. GDP and the mutated residues in Switch I were indicated.

In this paper, we have performed a range of MD simulations on wildtype K-Ras.GTP.GAP-334 and the E31N-Ras.GTP.GAP-334, D33N-Ras.GTP.GAP-334, D38N-Ras.GTP.GAP-334 and K-Ras.GTP.K935E-GAP-334 mutants, based on the MD trajectories we analysed structural conformation changes. We find that due to the mutations, the overall K-

55

Ras.GTP.GAP interaction for the mutant complexes becomes weaker compared to the wildtype system, of which the K935E mutation causes a most significant reduction in interaction energy, compare to other mutations. Mutation of negatively charged residues located in the Switch I region, namely E31N, D33N, D38N, or positively charged residue K935E of GAP spatially close to the Switch I region, influence the electrostatic interaction of the two proteins and thereby reduce their affinity to each other. Thus the complementary attractive electrostatic interaction is a key factor for the protein-protein interaction and recognition between K-Ras and GAP. By observing he distance between the atoms of the side chain phenol ring of Tyr32 in K-Ras and the atoms in the guanidinium group of catalytically critical Arg789 in GAP from different MD simulations of complexes, it was seen that mutations cause unfavorable conformations of the Switch I region and leave an incompletely open Tyr32 gate, which in turn hinders the interaction of the residue Arg789 from GAP to interact with GTP for the further acceleration of the GTP hydrolysis. Overall, this study enables us to better comprehend the precise changes in the protein structures that are brought out by mutated residues, and how this results in failed protein-protein interaction between K-Ras and GAP, and consequently leads to oncogenic effects.

# Chapter 6. Concluding remarks and future perspectives

CADD has become an essential part in the drug design process due to its powerful capability in the search of promising drug candidates. CADD makes use of the structural knowledge of either the target (structure-based) or known ligands with bioactivity (ligand-based) to facilitate every stage of the drug discovery and development process.

In this thesis, I first outline the mechanism or signalling pathway of protein targets, namely tyrosine kinase RET, mitosis kinesin Eg5 and KIF18B, histone acetyltransferase Tip60 and GTPase K-Ras, and illustrate how those targets are involved in developing cancer. Following that, I applied different CADD approaches to investigate protein structures, and protein-ligand interactions. The most widely used approaches in this thesis are homology modelling, docking, MD simulations, ligand-based and structure-based pharmacophore generation, as well as 3D-QSAR modelling. For the tyrosine kinase RET, we have 1) investigated the interaction between RET and DFG-out inhibitors and generated the key pharmacophore features of DFG-out inhibitors targeting RET; 2) studied the extracellular membrane domain of RET, the CLD region, to see how different mutations induce conformational changes. In addition, we proposed a pharmacophore model for targeting this extracellular region and screening for new inhibitors; 3) examined the possible mechanism of a photoswichable inhibitor targeting RET kinase active site. For the mitosis kinesin Eg5, we mainly focused on developing pharmacophore and 3D-QSAR model for a series of biphenyl type inhibitors, which target a recently found allosteric binding pocket of Eg5. In the case of the histone acetyltransferase Tip60, we collaborated with different groups, enabling rational design of a potential inhibitor, followed by its synthesis and validation of its activity in breast cancer cell lines. The GTPase K-Ras has been found to be mutated in different cancer types, and we herein studied several of its mutations located in the Switch I region to investigate how those mutations hinder the GTPase activity.

To be concluded, CADD is indeed a very useful tool for pharmaceutical companies and academic research groups to search for potential drug candidates with reduced cost and time. However, there is still room for further improvements in CADD, such as more accurate scoring functions, incorporating faster evaluation of target flexibility and solvent effects in the docking procedure, and increasing computational efficiency. To enhance current CADD, continuous improvements in the field of chemical and structural biology, bioinformatics, and computational technology is required. For the targets in this thesis there

is still room to continue developing inhibitors from different perspectives. For Ret, inhibitors e.g. can be developed to target the specific M918T mutant, which is mostly common in thyroid cancer; in addition, inhibitors targeting the extracellular domain of Ret are also promising to inhibit the oncogenic signalling. In the case of Eg5, inhibitors with new scaffolds and higher inhibitory activity targeting α4/α6 allosteric pocket could have significant impact. Finally, further optimization of the developed inhibitor TH1834 to increase its binding affinity for Tip60 is important in order to reach next stage of drug development.

# Acknowledgement

The whole PhD studies have been an incredible journey for me, it brings a great deal of memories and experiences. The most important thing I have gained professionally is not just the techniques I learned but also the scientific thinking I have developed. The most important thing I gain personally is I have had opportunities to meet and interact with many great people, whose talents in science and great personalities in private influence me significantly. Here, I would like to acknowledge some people who have been important to me during my PhD study period.

To my supervisor **Leif Eriksson**, thank you for taking me into your group from my master study until now, and always believing and supporting me during my whole study. Your great knowledge and expertise helped me incredibly to dig into my PhD projects. The personal experiences you shared with us inspired me a lot to be a better person. You are the best supervisor ever, I am honoured to have you as my supervisor.

To professor **Johan Gottfries**, many thanks for your excellent guidance in our SrtA projects and for always being willing to help me when I needed.

To professor **Morten Grotli**, thank you for being happy and willing to provide scientific guidance in our collaboration RET work, your enthusiasm in science inspired me a lot in my work.

To professor **Andrew Erwing**, thank you for being a good friend to discuss with me lots of useful information on working in science area, and provide me helpful information on raising a kid.

To all the co-authors of the papers, **Noel F. Lowndes, James A.L.Brown, Martin Scobie**, thank you for all the helpful discussions. It is nice to work with you.

To all friends and colleagues in the group and related groups in the past and present, **Anna, Antonio, Boxue, Li, Emma, Ivana, Martin, Min, Partricia, Patric, Yaoquan, Samuel, Thomas,** thank you for all the discussions in group meetings, and social group activities.

To **Eliza**, thank you for all those fun chats, and for always helpful when I need you.

To **Mozhgan**, thank you for your friendships and support all these years. You are the best friend that I always want to have. Thanks for all these phone calls, get-togethers and sharing with me your ups and downs in your life.

To **Bethmini and Nhu**, thank you for being my best friends after I came to Gotheburg. Thanks for all great memories we have had together: all the dinners, all the sharing moments, all the going out activities, all the gifts we gave to each other, my wedding, Daham and Estelle's birth, etc.

To my family, thanks to **Pengzhi** for being here with me, and support me with care and love. To our lovely daughter, **Estelle**, you are the little angel bringing me great joy and thank you for giving me the honor to be your mom.

# References

1. J. J. Manfredi and S. B. Horwitz, *Pharmacology & Therapeutics*, 1984, **25**, 83-125.
2. I. S. Johnson, J. G. Armstrong, M. Gorman and J. P. Burnett, *Cancer Research*, 1963, **23**, 1390-1427.
3. C. Tan, E. Etcubanas, N. Wollner, G. Rosen, A. Gilladoga, J. Showel, M. L. Murphy and I. H. Krakoff, *Cancer*, 1973, **32**, 9-17.
4. W. D. Kingsbury, J. C. Boehm, D. R. Jakas, K. G. Holden, S. M. Hecht, G. Gallagher, M. J. Caranfa, F. L. McCabe, L. F. Faucette and R. K. Johnson, *Journal of medicinal chemistry*, 1991, **34**, 98-107.
5. J. Zimmermann, E. Buchdunger, H. Mett, T. Meyer and N. B. Lydon, *Bioorganic & Medicinal Chemistry Letters*, 1997, **7**, 187-192.
6. T. O'Hare, D. K. Walters, M. W. Deininger and B. J. Druker, *Cancer cell*, 2005, **7**, 117-119.
7. T. Kim and J. Murren, *Current opinion in investigational drugs (London, England: 2000)*, 2002, **3**, 1385-1395.
8. N. Howlader, A. Noone, M. Krapcho, N. Neyman, R. Aminou, S. Altekruse, C. Kosary, J. Ruhl, Z. Tatalovich and H. Cho, *Journal*, 2012.
9. E. A. Gombos, E. D. Freis and A. Moghadam, *New England Journal of Medicine*, 1966, **275**, 1215-1220.
10. D. Campbell and E. Phillips, *European journal of clinical pharmacology*, 1974, **7**, 407-414.
11. P. Roy, L. Day and E. Sowton, *Br Med J*, 1975, **3**, 195-197.
12. R. Shanks, T. Wood, A. Dornhorst and M. Clark, 1966.
13. M. A. Ondetti, B. Rubin and D. W. Cushman, *Science*, 1977, **196**, 441-444.
14. E. S. Istvan and J. Deisenhofer, *Science*, 2001, **292**, 1160-1164.
15. C. Cannon, *Diabetes*, 1911, **91**, 90.94.
16. P. S. Collaboration, *The Lancet*, 2007, **370**, 1829-1839.
17. I. M. Kapetanovic, *Chemico-Biological Interactions*, 2008, **171**, 165-176.
18. I. Kapetanovic, *Chemico-biological interactions*, 2008, **171**, 165-176.
19. R. Siegel, D. Naishadham and A. Jemal, *CA: a cancer journal for clinicians*, 2013, **63**, 11-30.
20. I. J. Fidler, *Nature Reviews Cancer*, 2003, **3**, 453-458.
21. I. Kola and J. Landis, *Nature reviews Drug discovery*, 2004, **3**, 711-716.
22. T. G. Roberts, B. H. Goulart, L. Squitieri, S. C. Stallings, E. F. Halpern, B. A. Chabner, G. S. Gazelle, S. N. Finkelstein and J. W. Clark, *Jama*, 2004, **292**, 2130-2140.
23. C. Sawyers, *Nature*, 2004, **432**, 294-297.
24. I. Collins and P. Workman, *Nature Chemical Biology*, 2006, **2**, 689-700.
25. A. Kamb, S. Wee and C. Lengauer, *Nat Rev Drug Discov*, 2007, **6**, 115-120.
26. N. Kumar, B. S. Hendriks, K. A. Janes, D. de Graaf and D. A. Lauffenburger, *Drug discovery today*, 2006, **11**, 806-811.
27. J. H. Van Drie, *Journal of computer-aided molecular design*, 2007, **21**, 591-601.
28. G. Sliwoski, S. Kothiwale, J. Meiler and E. W. Lowe, *Pharmacological Reviews*, 2014, **66**, 334-395.
29. R. Vijayakrishnan, *Journal of postgraduate medicine*, 2009, **55**, 301.
30. T. T. Talele, S. A. Khedkar and A. C. Rigby, *Current topics in medicinal chemistry*, 2010, **10**, 127-141.

31. A. H. Gradman, R. E. Schmieder, R. L. Lins, J. Nussberger, Y. Chiang and M. P. Bedigian, *Circulation*, 2005, **111**, 1012-1018.
32. J. Singh, C. E. Chuaqui, P. A. Boriack-Sjodin, W.-C. Lee, T. Pontz, M. J. Corbley, H.-K. Cheung, R. M. Arduini, J. N. Mead and M. N. Newman, *Bioorganic & medicinal chemistry letters*, 2003, **13**, 4355-4359.
33. J. S. Sawyer, B. D. Anderson, D. W. Beight, R. M. Campbell, M. L. Jones, D. K. Herron, J. W. Lampe, J. R. McCowan, W. T. McMillen, N. Mort, S. Parsons, E. C. R. Smith, M. Vieth, L. C. Weir, L. Yan, F. Zhang and J. M. Yingling, *Journal of Medicinal Chemistry*, 2003, **46**, 3953-3956.
34. Q. Vadis, *Journal of Medicinal Chemistry*, 2010, **53**, 8461-8467.
35. D. V. S. Green, in *Progress in Medicinal Chemistry*, Elsevier, 2003, vol. Volume 41, pp. 61-97.
36. E. Bharath, S. Manjula and A. Vijaychand, *Chemistry*, 2011, **18**, 1.0.
37. K. M. Merz Jr, D. Ringe and C. H. Reynolds, *Drug design: structure-and ligand-based approaches*, Cambridge University Press, 2010.
38. T. I. Oprea and H. Matter, *Current opinion in chemical biology*, 2004, **8**, 349-358.
39. D. B. Kitchen, H. Decornez, J. R. Furr and J. Bajorath, *Nature reviews Drug discovery*, 2004, **3**, 935-949.
40. G. Klebe, *Drug discovery today*, 2006, **11**, 580-594.
41. O. Dror, A. Shulman-Peleg, R. Nussinov and H. J. Wolfson, *Current medicinal chemistry*, 2004, **11**, 71-90.
42. E. Kellenberger, J. Rodrigo, P. Muller and D. Rognan, *Proteins: Structure, Function, and Bioinformatics*, 2004, **57**, 225-242.
43. E. Perola, *Proteins: Structure, Function, and Bioinformatics*, 2006, **64**, 422-435.
44. H. Steuber, M. Zentgraf, C. Gerlach, C. A. Sotriffer, A. Heine and G. Klebe, *Journal of molecular biology*, 2006, **363**, 174-187.
45. G. Scapin, *Current pharmaceutical design*, 2006, **12**, 2087-2097.
46. A. R. Leach, B. K. Shoichet and C. E. Peishoff, *Journal of medicinal chemistry*, 2006, **49**, 5851-5855.
47. L. M. Mulligan, *Nature Reviews Cancer*, 2014, **14**, 173-186.
48. M. S. Airaksinen, A. Titievsky and M. Saarma, *Molecular and Cellular Neuroscience*, 1999, **13**, 313-325.
49. H. Murakami, T. Iwashita, N. Asai, Y. Shimono, Y. Iwata, K. Kawai and M. Takahashi, *Biochemical and biophysical research communications*, 1999, **262**, 68-75.
50. C. Segouffin-Cariou and M. Billaud, *Journal of Biological Chemistry*, 2000, **275**, 3568-3576.
51. M. Chiariello, R. Visconti, F. Carlomagno, R. M. Melillo, C. Bucci, V. de Franciscis, G. M. Fox, S. Jing, O. A. Coso and J. S. Gutkind, *Oncogene*, 1998, **16**, 2435-2445.
52. E. Arighi, M. G. Borrello and H. Sariola, *Cytokine & growth factor reviews*, 2005, **16**, 441-467.
53. T. S. Gujral, W. Van Veelen, D. S. Richardson, S. M. Myers, J. A. Meens, D. S. Acton, M. Duñach, B. E. Elliott, J. W. Höppener and L. M. Mulligan, *Cancer Research*, 2008, **68**, 1338-1346.
54. H. Prazeres, J. Torres, F. Rodrigues, J. P. Couto, J. Vinagre, #xe3, M. Sobrinho-Simões and P. Soares, *Journal of Thyroid Research*, 2011, **2011**, 10.
55. J. W. B. de Groot, T. P. Links, J. T. Plukker, C. J. Lips and R. M. Hofstra, *Endocrine reviews*, 2006, **27**, 535-560.
56. G. Cuccuru, C. Lanzi, G. Cassinelli, G. Pratesi, M. Tortoreto, G. Petrangolini, E. Seregni, A. Martinetti, D. Laccabue and C. Zanchi, *Journal of the National Cancer Institute*, 2004, **96**, 1006-1014.

57. C. Lanzi, G. Cassinelli, T. Pensa, M. Cassinis, R. A. Gambetta, M. G. Borrello, E. Menta, M. A. Pierotti and F. Zunino, *International Journal of Cancer*, 2000, **85**, 384-390.

58. M. S. Cohen, H. B. Hussain and J. F. Moley, *Surgery*, 2002, **132**, 960-967.

59. S. Ezzat, P. Huang, A. Dackiw and S. L. Asa, *Clinical cancer research*, 2005, **11**, 1336-1341.

60. C. J. Strock, J.-I. Park, M. Rosen, C. Dionne, B. Ruggeri, S. Jones-Bolin, S. R. Denmeade, D. W. Ball and B. D. Nelkin, *Cancer Research*, 2003, **63**, 5559-5563.

61. F. Carlomagno, D. Vitagliano, T. Guida, M. Napolitano, G. Vecchio, A. Fusco, A. Gazit, A. Levitzki and M. Santoro, *Cancer Research*, 2002, **62**, 1077-1082.

62. C. Carniti, C. Perego, P. Mondellini, M. A. Pierotti and I. Bongarzone, *Cancer research*, 2003, **63**, 2234-2243.

63. F. Carlomagno, D. Vitagliano, T. Guida, F. Ciardiello, G. Tortora, G. Vecchio, A. J. Ryan, G. Fontanini, A. Fusco and M. Santoro, *Cancer research*, 2002, **62**, 7284-7290.

64. P. J. Horn and C. L. Peterson, *Science*, 2002, **297**, 1824-1827.

65. A. Annunziato, *Nature Education*, 2008, **1**, 26.

66. A. Vaquero, A. Loyola and D. Reinberg, *Science's SAGE KE*, 2003, **2003**, 4.

67. S. Y. Roth, J. M. Denu and C. D. Allis, *Annual review of biochemistry*, 2001, **70**, 81-120.

68. N. Avvakumov and J. Cote, *Oncogene*, 2007, **26**, 5395-5407.

69. D. Phillips, *Biochemical Journal*, 1963, **87**, 258.

70. V. Sapountzi, I. R. Logan and C. N. Robson, *The International Journal of Biochemistry & Cell Biology*, 2006, **38**, 1496-1509.

71. B. J. Feldman and D. Feldman, *Nature Reviews Cancer*, 2001, **1**, 34-45.

72. L. Gaughan, I. R. Logan, S. Cook, D. E. Neal and C. N. Robson, *Journal of Biological Chemistry*, 2002, **277**, 25904-25913.

73. J. H. Patel, Y. Du, P. G. Ard, C. Phillips, B. Carella, C.-J. Chen, C. Rakowski, C. Chatterjee, P. M. Lieberman and W. S. Lane, *Molecular and cellular biology*, 2004, **24**, 10826-10834.

74. W. Fischle, *Nature cell biology*, 2009, **11**, 1279-1281.

75. M. Squatrito, C. Gorrini and B. Amati, *Trends in cell biology*, 2006, **16**, 433-442.

76. Y. Tang, J. Luo, W. Zhang and W. Gu, *Molecular cell*, 2006, **24**, 827-839.

77. Y. Sun, X. Jiang, S. Chen, N. Fernandes and B. D. Price, *Proceedings of the National Academy of Sciences of the United States of America*, 2005, **102**, 13182-13187.

78. G. Nunnari, J. A. Smith and R. Daniel, *Journal of Experimental & Clinical Cancer Research*, 2008, **27**, 1-8.

79. Y. Sun, X. Jiang, S. Chen and B. D. Price, *FEBS letters*, 2006, **580**, 4353-4356.

80. D. J. Brenner, R. Doll, D. T. Goodhead, E. J. Hall, C. E. Land, J. B. Little, J. H. Lubin, D. L. Preston, R. J. Preston and J. S. Puskin, *Proceedings of the National Academy of Sciences*, 2003, **100**, 13761-13766.

81. O. D. Lau, T. K. Kundu, R. E. Soccio, S. Ait-Si-Ali, E. M. Khalil, A. Vassilev, A. P. Wolffe, Y. Nakatani, R. G. Roeder and P. A. Cole, *Molecular cell*, 2000, **5**, 589-595.

82. T. Oike, H. Ogiwara, K. Torikai, T. Nakano, J. Yokota and T. Kohno, *International Journal of Radiation Oncology* Biology* Physics*, 2012, **84**, 815-821.

83. F. J. Dekker, M. Ghizzoni, N. van der Meer, R. Wisastra and H. J. Haisma, *Bioorganic & medicinal chemistry*, 2009, **17**, 460-466.

84. S. Gorsuch, V. Bavetsias, M. G. Rowlands, G. W. Aherne, P. Workman, M. Jarman and E. McDonald, *Bioorganic & medicinal chemistry*, 2009, **17**, 467-474.

85. M. Biel, A. Kretsovali, E. Karatzali, J. Papamatheakis and A. Giannis, *Angewandte Chemie International Edition*, 2004, **43**, 3974-3976.

86. A. Mai, D. Cheng, M. T. Bedford, S. Valente, A. Nebbioso, A. Perrone, G. Brosch, G. Sbardella, F. De Bellis and M. Miceli, *Journal of medicinal chemistry*, 2008, **51**, 2279-2290.
87. G. Sbardella, S. Castellano, C. Vicidomini, D. Rotili, A. Nebbioso, M. Miceli, L. Altucci and A. Mai, *Bioorganic & medicinal chemistry letters*, 2008, **18**, 2788-2792.
88. D. A. Compton, *Annual review of biochemistry*, 2000, **69**, 95-114.
89. S. Inoué and E. D. Salmon, *Molecular Biology of the Cell*, 1995, **6**, 1619-1640.
90. M. A. Jordan, D. Thrower and L. Wilson, *Cancer research*, 1991, **51**, 2212-2222.
91. E. K. Rowinsky, N. Onetto, R. M. Canetta and S. G. Arbuck, 1992.
92. M. A. Jordan and L. Wilson, *Nature Reviews Cancer*, 2004, **4**, 253-265.
93. P. Dustin, *Microtubules*, Springer Science & Business Media, 2012.
94. H. Miki, Y. Okada and N. Hirokawa, *Trends in cell biology*, 2005, **15**, 467-476.
95. V. Mountain and D. A. Compton, *The Anatomical Record*, 2000, **261**, 14-24.
96. E. Mandelkow and E.-M. Mandelkow, *Trends in Cell Biology*, 2002, **12**, 585-591.
97. M. T. Valentine, P. M. Fordyce and S. M. Block, *Cell Division*, 2006, **1**, 31-31.
98. D. Huszar, M.-E. Theoclitou, J. Skolnik and R. Herbst, *Cancer and Metastasis Reviews*, 2009, **28**, 197-208.
99. H. B. El-Nassan, *European Journal of Medicinal Chemistry*, 2013, **62**, 614-631.
100. P. Hegde, J. Cogswell, K. Carrick, J. Jackson, K. Wood, W. Eng, M. Brawner, P. Huang and D. Bergsma, 2003.
101. B. Z. Carter, D. H. Mak, Y. Shi, W. D. Schober, R.-Y. Wang, M. Konopleva, E. Koller, N. M. Dean and M. Andreeff, *Cell Cycle*, 2006, **5**, 2223-2229.
102. S. Ding, N. Xing, J. Lu, H. Zhang, K. Nishizawa, S. Liu, X. Yuan, Y. Qin, Y. Liu and O. Ogawa, *International Journal of Urology*, 2011, **18**, 432-438.
103. M. Liu, X. Wang, Y. Yang, D. Li, H. Ren, Q. Zhu, Q. Chen, S. Han, J. Hao and J. Zhou, *The Journal of pathology*, 2010, **221**, 221-228.
104. V. Sarli and A. Giannis, *ChemMedChem*, 2006, **1**, 293-298.
105. R. Sakowicz, J. T. Finer, C. Beraud, A. Crompton, E. Lewis, A. Fritsch, Y. Lee, J. Mak, R. Moody and R. Turincio, *Cancer research*, 2004, **64**, 3276-3280.
106. H. Yokoyama, J.-i. Sawada, S. Katoh, K. Matsuno, N. Ogo, Y. Ishikawa, H. Hashimoto, S. Fujii and A. Asai, *ACS chemical biology*, 2015, **10**, 1128-1136.
107. M. E. Tanenbaum, L. Macurek, B. van der Vaart, M. Galli, A. Akhmanova and R. H. Medema, *Current Biology*, 2011, **21**, 1356-1365.
108. J. R. Stout, A. L. Yount, J. A. Powers, C. LeBlanc, S. C. Ems-McClung and C. E. Walczak, *Molecular biology of the cell*, 2011, **22**, 3070-3080.
109. T. Itzel, P. Scholz, T. Maass, M. Krupp, J. U. Marquardt, S. Strand, D. Becker, F. Staib, H. Binder and S. Roessler, *Bioinformatics*, 2014, btu586.
110. L. Frizzell, 2015.
111. K. Wennerberg, K. L. Rossman and C. J. Der, *Journal of Cell Science*, 2005, **118**, 843-846.
112. K. Wennerberg, K. L. Rossman and C. J. Der, *J Cell Sci*, 2005, **118**, 843-846.
113. S. A. Forbes, N. Bindal, S. Bamford, C. Cole, C. Y. Kok, D. Beare, M. Jia, R. Shepherd, K. Leung, A. Menzies, J. W. Teague, P. J. Campbell, M. R. Stratton and P. A. Futreal, *Nucleic Acids Research*, 2011, **39**, D945-D950.
114. I. A. Prior, P. D. Lewis and C. Mattos, *Cancer Research*, 2012, **72**, 2457-2467.
115. Y. Wang, C. E. Kaiser, B. Frett and H.-y. Li, *Journal of medicinal chemistry*, 2013, **56**, 5219-5230.
116. I. R. Vetter and A. Wittinghofer, *Science*, 2001, **294**, 1299-1304.
117. J. Cherfils and M. Zeghouf, *Physiological reviews*, 2013, **93**, 269-309.

118. A. M. Rojas, G. Fuentes, A. Rausell and A. Valencia, *The Journal of cell biology*, 2012, **196**, 189-201.
119. T. Maurer, L. S. Garrenton, A. Oh, K. Pitts, D. J. Anderson, N. J. Skelton, B. P. Fauber, B. Pan, S. Malek and D. Stokoe, *Proceedings of the National Academy of Sciences*, 2012, **109**, 5299-5304.
120. N. Vasan, J. L. Boyer and R. S. Herbst, *Clinical Cancer Research*, 2014, **20**, 3921-3930.
121. Q. Sun, J. P. Burke, J. Phan, M. C. Burns, E. T. Olejniczak, A. G. Waterson, T. Lee, O. W. Rossanese and S. W. Fesik, *Angewandte Chemie (International ed. in English)*, 2012, **51**, 6140-6143.
122. J. M. Ostrem, U. Peters, M. L. Sos, J. A. Wells and K. M. Shokat, *Nature*, 2013, **503**, 548-551.
123. C. Chothia, *Nature*, 1992, **357**, 543-544.
124. A. Hillisch, L. F. Pineda and R. Hilgenfeld, *Drug discovery today*, 2004, **9**, 659-669.
125. S. F. Altschul, W. Gish, W. Miller, E. W. Myers and D. J. Lipman, *Journal of Molecular Biology*, 1990, **215**, 403-410.
126. W. R. Pearson, in *Methods in Enzymology*, Academic Press, 1990, vol. Volume 183, pp. 63-98.
127. C. N. Cavasotto and S. S. Phatak, *Drug discovery today*, 2009, **14**, 676-683.
128. L. Prime. Schrödinger, Portland, OR.
129. T. Schwede, J. Kopp, N. Guex and M. C. Peitsch, *Nucleic acids research*, 2003, **31**, 3381-3385.
130. M. MOE. Chemical Computing Group and C. Quebec.
131. A. Sali and T. Blundell, *Protein structure by distance analysis*, 1994, **64**, C86.
132. K. M. Misura, D. Chivian, C. A. Rohl, D. E. Kim and D. Baker, *Proceedings of the National Academy of Sciences*, 2006, **103**, 5361-5366.
133. M. A. Dolan, J. W. Noah and D. Hurt, *Homology Modeling: Methods and Protocols*, 2012, 399-414.
134. I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge and T. E. Ferrin, *Journal of Molecular Biology*, 1982, **161**, 269-288.
135. D. B. Kitchen, H. Decornez, J. R. Furr and J. Bajorath, *Nat Rev Drug Discov*, 2004, **3**, 935-949.
136. T. Darden, D. York and L. Pedersen, *The Journal of Chemical Physics*, 1993, **98**, 10089-10092.
137. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *The Journal of Chemical Physics*, 1995, **103**, 8577-8593.
138. R. W. Hockney, *POTENTIAL CALCULATION AND SOME APPLICATIONS*, Langley Research Center, Hampton, Va., 1970.
139. L. Verlet, *Physical review*, 1967, **159**, 98.
140. G. Wolber and T. Langer, *Journal of Chemical Information and Modeling*, 2005, **45**, 160-169.
141. F. Ortuso, T. Langer and S. Alcaro, *Bioinformatics*, 2006, **22**, 1449-1455.
142. J. Chen and L. Lai, *Journal of Chemical Information and Modeling*, 2006, **46**, 2684-2691.
143. H. Yang, Y. Shen, J. Chen, Q. Jiang, Y. Leng and J. Shen, *European Journal of Medicinal Chemistry*, 2009, **44**, 1167-1171.
144. R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin and D. T. Mainz, *Journal of Medicinal Chemistry*, 2006, **49**, 6177-6196.

145. N. K. Salam, R. Nuti and W. Sherman, *Journal of Chemical Information and Modeling*, 2009, **49**, 2356-2368.
146. H. Wang, R. A. Duffy, G. C. Boykow, S. Chackalamannil and V. S. Madison, *Journal of Medicinal Chemistry*, 2008, **51**, 2439-2446.
147. D. Schuster, L. G. Nashev, J. Kirchmair, C. Laggner, G. Wolber, T. Langer and A. Odermatt, *Journal of Medicinal Chemistry*, 2008, **51**, 4188-4199.
148. M. A. C. Neves, T. C. P. Dinis, G. Colombo and M. L. Sá e Melo, *Journal of Medicinal Chemistry*, 2009, **52**, 143-150.
149. D. A. Brown, P. S. Kharkar, I. Parrington, M. E. A. Reith and A. K. Dutta, *Journal of Medicinal Chemistry*, 2008, **51**, 7806-7819.
150. S. L. Dixon, A. M. Smondyrev, E. H. Knoll, S. N. Rao, D. E. Shaw and R. A. Friesner, *Journal of Computer-Aided Molecular Design*, 2006, **20**, 647-671.
151. G. Chang, W. C. Guida and W. C. Still, *Journal of the American Chemical Society*, 1989, **111**, 4379-4386.
152. I. Kolossváry and W. C. Guida, *Journal of the American Chemical Society*, 1996, **118**, 5011-5019.
153. W. Koch, M. C. Holthausen and M. Kaupp, *Angewandte Chemie-German Edition*, 2001, **113**, 989-989.
154. R. Ferreira, J. R. Nilsson, C. Solano, J. Andréasson and M. Grøtli, *Scientific reports*, 2015, **5**.