

---

**IN-SILICO INVESTIGATION OF COENZYME A SELECTIVITY FOR  
AURORA A KINASE AND DEVELOPMENT OF AN AURORA A KINASE-  
SELECTIVE INHIBITOR AS A POTENTIAL ANTICANCER AGENT.**

---

Trang M. Tran

Ph.D

Supervisor: Prof. Alethea B. Tabor

Submitted in partial fulfilment of the requirements of the Doctor of Philosophy in Chemistry of  
University College London.

NOVEMBER 1, 2016

UCL

**Declaration**

*I, Trang M. Tran, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.*

London, 1<sup>st</sup> May, 2016

*Trang M Tran*

## Abstract

Centrosome amplification has been observed in most cancer cells, and is considered to be a “hallmark” of cancer cells. This process is commonly associated with chromosome segregation process in the mitosis phase in the cell cycle, which is tightly controlled by mitotic kinases. Among these kinases, the Aurora kinase family, Aurora A (AURKA), Aurora B (AURKB) and Aurora C (AURKC), ensures the accurate progression of mitosis, including the formation of a bipolar mitotic spindle, accurate segregation of chromosomes and the completion of cytokinesis. AURKA has been seen to have the largest role in mitotic progression and checkpoint control pathways and overexpression of AURKA is most associated with cancer. Hence, interfering with AURKA activity has been considered to be a promising approach to anticancer agents. Professor Gout’s group has recently shown that Coenzyme A (CoA) selectively inhibits AURKA activity ( $IC_{50}$  of 4.4  $\mu$ M). However, its large molecular weight (>500) and negatively charged phosphate group make it unsuitable as a drug candidate. This project was set to investigate the possible binding modes of CoA in the catalytic domain of AURKA. The corresponding interactions between CoA and protein residues would provide some insights in the selectivity of CoA towards AURKA. Furthermore, based on the understanding of the interactions between CoA and the catalytic domain of AURKA and the possible reasons behind the selectivity of CoA towards AURKA, *in silico* design and synthesis of a new highly selective and potent AURKA inhibitor based on the structure of CoA and current lead compounds which are in clinical trials for Aurora kinase inhibitors could be carried out.

## **Acknowledgement**

I would like to thank my supervisor, Prof. Alethea B Tabor and Prof. Ivan Gout for the opportunity to work on this project.

I would like to thank Dr. Edith Chan for her guidance and support in computational study.

I would like to thank Dr. Tsuchiya for his support in biological images and data.

I would like to thank Ms. Fiona Bellany for her collaboration in synthetic chemistry throughout my research and Ms. Yueyang Huang for her assistance in pharmacological testings. I would like to thank Dr. Richard Bayliss and Dr. Patrick Eysers and Mr. Patrick McIntyre – for their collaborative work in crystallography. I would like to thank Dr. Lalitha Guruprasad for the preliminary docking studies. I would also like to gratefully acknowledge Dr. Kersti Karu and Dr. Lisa D Haigh's help with running mass spectroscopy samples, Dr. Abil E Aliev for his help with NMR analysis and members of Tabor's and Hailes' research groups for their help and advice.

My gratitude also goes to all the past and present members who have been working on the projects.

**Funding.** I would like to express my gratitude to UCL for the generous fundings of UCL Graduate Research Scholarship and UCL Overseas Research Scholarship.

## TABLE OF CONTENTS

Declaration.....	1
Abstract.....	2
Acknowledgement .....	3
Abbreviations.....	6
<b>1. Introduction of biochemistry background and thesis proposal.....</b>	<b>8</b>
1.1. An introduction on cancer .....	8
1.2. Protein kinases for cancer therapy .....	15
1.3. Aurora A kinase (AURKA) for cancer therapy .....	25
1.4. Biochemical findings .....	38
1.5. Thesis proposal .....	43
1.5.1. Approach A – The first generation of compounds .....	44
1.5.2. Approach B – The second generation of compounds.....	46
<b>2. Computational study and design .....</b>	<b>47</b>
2.1. Backgrounds on proteins and methods of modelling.....	47
2.1.1. Backgrounds .....	47
2.1.2. Materials and methods of modelling.....	54
2.2. Backgrounds on computational tools.....	62
2.2.1. PROCHECK for Ramachandran plot .....	62
2.2.2. MOE for molecular visualisation and amino acid alignment and modification	62
2.2.3. Other bioinformatics tools.....	62
2.3. Results and Discussion on the selectivity of CoA towards AURKA.....	63
2.3.1. Structural alignment of selected protein kinases for structural comparison.....	63
2.3.2. Docking results and discussion of the conformation of CoA in AURKA .....	72
2.4. Ab initio design of a new AURKA selective inhibitor.....	90
2.4.1. Re-examination of preliminary docking results in the 1 <sup>st</sup> generation of compounds .....	90
2.4.2. Methods of new design for the 2 <sup>nd</sup> generation of compounds .....	93

2.5.	Conclusion .....	105
<b>3.</b>	<b>Approach A syntheses.....</b>	<b>107</b>
3.1.	Synthesis of the first generation of compounds .....	107
3.1.1.	<i>Literature research and synthetic plan of the 1<sup>st</sup> generation of compounds .....</i>	<i>107</i>
3.1.2.	<i>Synthetic results of pantothenamide tail .....</i>	<i>110</i>
3.1.3.	<i>Synthesis of linkers and esterification.....</i>	<i>111</i>
3.2.	Heteroaromatic core.....	115
3.2.1.	<i>Literature and preliminary research on synthesis of the heteroaromatic head group, the first generation of compounds .....</i>	<i>115</i>
3.2.2.	<i>Synthetic results .....</i>	<i>119</i>
3.3.	Conclusion .....	123
<b>4.</b>	<b>Approach B syntheses.....</b>	<b>124</b>
4.1.	Literature research and proposed approach.....	125
4.2.	Synthetic results .....	127
4.2.1.	<i>The feasibility of forming prerequisite alkynes and azides .....</i>	<i>127</i>
4.2.2.	<i>The feasibility of Click reaction between alkyne 72 and azide 64 .....</i>	<i>130</i>
4.3.	Biological screening results of the second generation of compounds and Discussion ... .....	137
<b>5.</b>	<b>Future work.....</b>	<b>138</b>
<b>6.</b>	<b>Synthetic experimental .....</b>	<b>139</b>
6.1.	Approach A syntheses.....	140
6.1.1.	<i>Synthesis of the first generation of compounds .....</i>	<i>140</i>
6.1.2.	<i>Heteroaromatic core .....</i>	<i>150</i>
6.2.	Approach B syntheses.....	159
	<i>Failed synthetic attempts .....</i>	<i>178</i>
<b>7.</b>	<b>Appendix.....</b>	<b>187</b>
<b>8.</b>	<b>References.....</b>	<b>196</b>

## Abbreviations

ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
HOAt	1-Hydroxy-7-azabenzotriazole
Boc <sub>2</sub> O	Di- <i>tert</i> -butyl dicarbonate (Boc anhydride)
BzCl	Benzoyl chloride
CoA	Coenzyme A
DavePhos	Dicyclohexylphosphino-2'-( <i>N,N</i> -dimethylamino)biphenyl
DCC	<i>N,N'</i> -dicyclohexylcarbodiimide
DIPEA	Diisopropylethylamine
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide
DPPA	Diphenylphosphoryl azide
EtOAc	Ethyl acetate
Fmoc	Fluorenylmethoxycarbonyl chloride
FmocOSu	<i>N</i> -(9-Fluorenylmethoxycarbonyloxy)succinimide
Glu	Glutamic acid
HBTU	<i>O</i> -Benzotriazole- <i>N,N,N',N'</i> -tetramethyl-uronium-hexafluoro-phosphate
HATU	(1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate)
HRMS	High resolution mass spectroscopy
HPLC	High performance liquid chromatography
Hex	Hexane
IC <sub>50</sub>	Half maximal inhibitory concentration
LCMS	Liquid chromatography mass spectrometry
micrw	Microwave
MHz	Megahertz
MW	Molecular weight
NHEt <sub>2</sub>	Diethylamine
NH <sub>4</sub> OAc	Ammonium acetate
NMM	<i>N</i> -methylmorpholine
PMBCl	<i>para</i> -methoxybenzyl chloride
PyAOP	((7-Azabenzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate)
QPhos	1,2,3,4,5-Pentaphenyl-1'-(di- <i>tert</i> -butylphosphino)ferrocene
TBAI	Tetrabutylammonium iodide
( <i>R</i> )-BINAP	( <i>R</i> )-(+)-(1,1'-Binaphthalene-2,2'-diyl)bis(diphenylphosphine)

<i>t</i> -Bu	<i>tert</i> -Butyl
<i>t</i> -BuXPhos	2-Di- <i>tert</i> -butylphosphino-2',4',6'-triisopropylbiphenyl
TBTA	Tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine
TCEP	Tris-(carboxyethyl) phosphine hydrochloride
TEA or NEt <sub>3</sub>	Triethylamine
THF	Tetrahydrofuran
THPTA	Tris(3-hydroxypropyltriazolylmethyl)amine
TFA	Trifluoroacetic acid
TLC	Thin layer chromatography
TMSCl	Trimethylsilyl chloride



## 1. Introduction of biochemistry background and thesis proposal

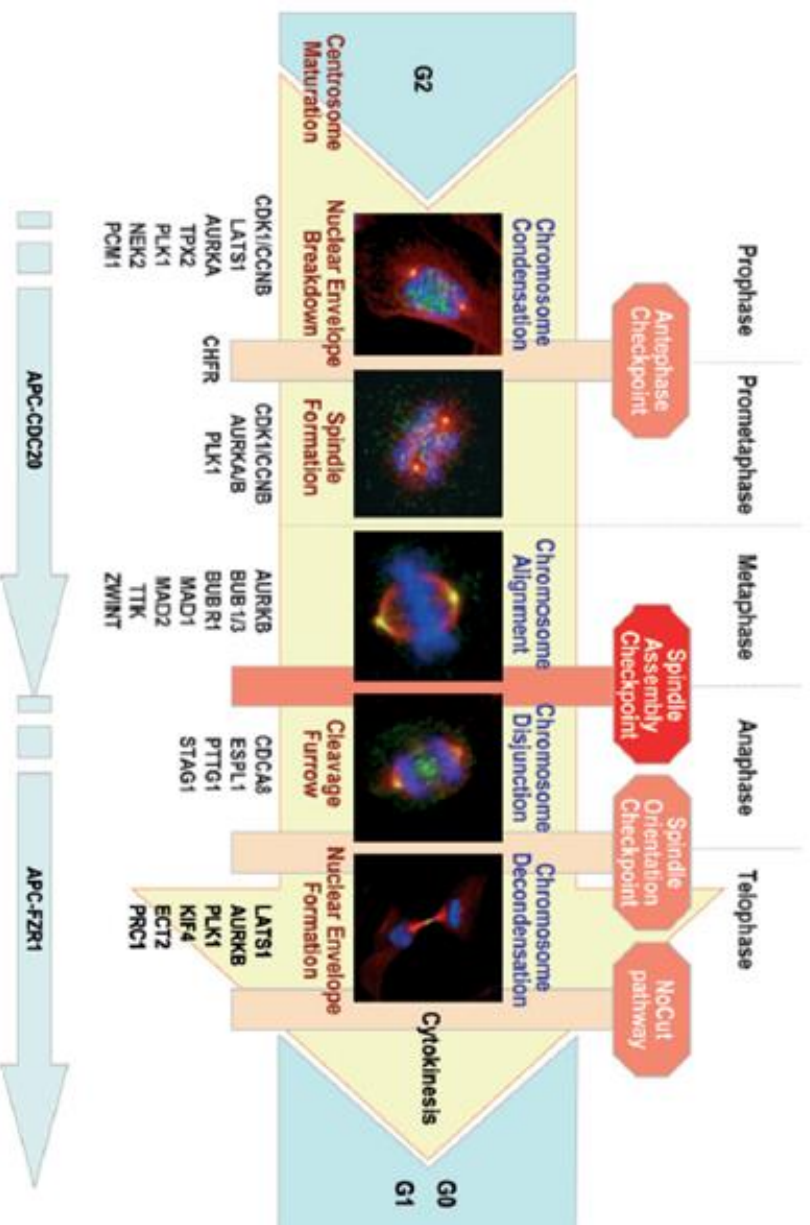
### 1.1. An introduction on cancer <sup>1</sup>

Cancer, known otherwise as malignant tumours and neoplasms, is a group of diseases that can affect any part of the body, causing 15% of deaths worldwide (8.2 million) in 2012. <sup>2</sup> In the UK, approximately 1 in 4 of all deaths in 2011 were due to cancer. In 2014, there were 163,444 death counts in the UK with cancer being listed as the cause. Almost half of all the deaths due to cancer were lung, bowel, breast or prostate (UK, 2014). <sup>2</sup> The risk of developing cancer in humans depends on many factors, including age, genetics and environmental factors such as lifestyle (e.g. diet, smoking) and exposure to chemicals, radiation and pollution. According to Cancer Research UK, 42% of cancer cases in the UK each year are linked to lifestyle factors. In particular, smoking is the largest single cause of cancer in the UK. It accounts for approximately 19% of cancer cases each year in the UK, with lung cancer having the highest proportion of smoking-linked cases. Worldwide, tobacco is also considered the single most important risk factor for cancer, linked to 71% of lung cancer deaths in 2008 (approximately 1 million).

The main features of cancer include uncontrolled growth and proliferation of cells followed by loss of function and de-differentiation of cells forming localised tumours. The localised tumours can then invade the surroundings and metastasise to other parts of the body through the bloodstream or lymphatic system, forming a secondary (metastatic) deposit. <sup>1c</sup> In order to ensure the nourishment for the malignant cells to survive and grow, formation of new blood vessels occurs (angiogenesis).

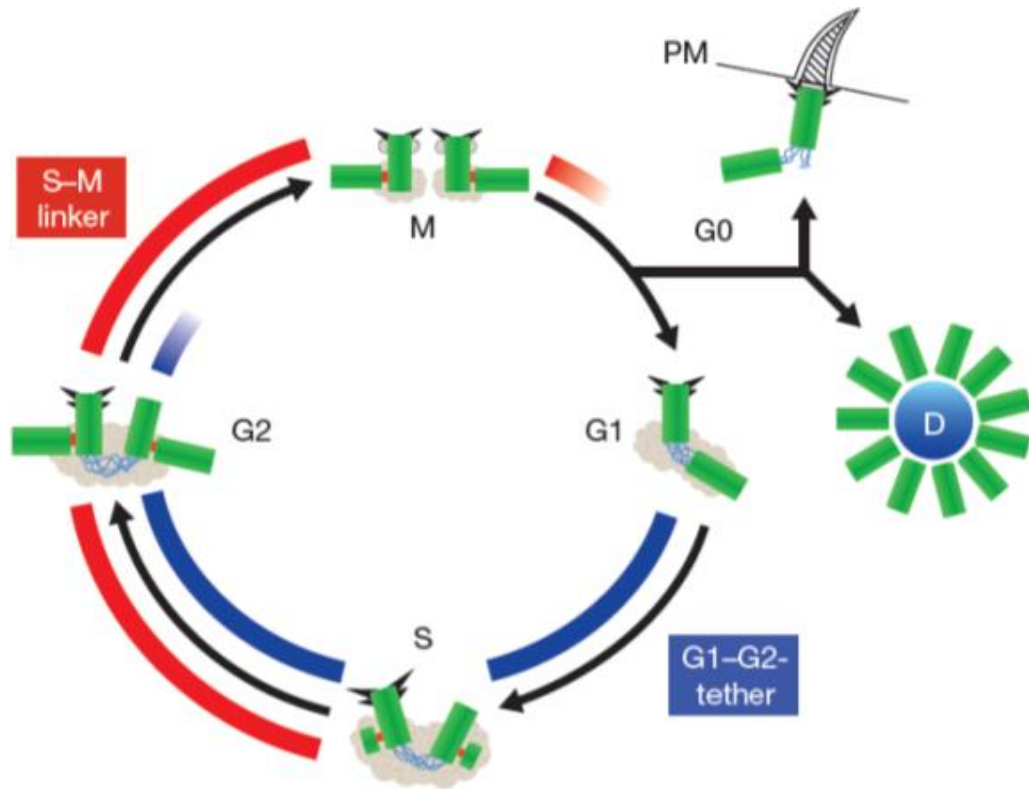
At the molecular level, the causes of cancer lie in the breakdown of the mechanisms regulating cell division and cell proliferation (mitosis). A cell is the fundamental unit of all living matter. There are two types of cells: prokaryotes (cells that lack a nuclear membrane) and eukaryotes (cells that contain membrane-bound cell organelles). A human cell is eukaryotic, which means it contains a true nucleus. A nucleus has three main components: a nucleolus, which produces ribosomes, a nuclear envelope, which is a double-layered membrane separating the nucleus from the cytoplasm and chromatin, and a DNA/protein complex containing our genes. All the cells in a human body are classified as belonging to either the germ-line or somatic tissue. Germ-line tissue is responsible for the formation of the gonads. All other cells of the body belong to somatic tissues. Most cancers arise in somatic rather than germ cells. <sup>1a, b</sup> An active dividing eukaryotic cell (somatic cell) passes through a series of stages known as *the cell cycle*. The cell cycle consists of four phases, G1, S, G2 and M phase. The G1 phase is where the cell grows in size and metabolic changes occur to prepare for cell division. At the G1 stage, cells may leave the cell cycle for a resting period (G0). Otherwise, after the G1 phase, cells may

enter S phase where DNA synthesis occurs, leading to the replication of genetic materials. The centrosome, an organelle near the nucleus that is responsible for organising microtubules throughout the cell cycle, begins its duplication process. The G2 phase is where metabolic changes prepare the cytoplasmic materials for mitosis and cytokinesis. The M phase is where nuclear division (mitosis) occurs followed by cell division (cytokinesis). Mitosis is further divided into 5 phases. The prophase is the first phase of mitosis where the genetic material, duplicated chromatin, condenses into visible chromosomes. Each chromosome contains two genetically identical sister chromatids, joined together at the centromere. Microtubules, which construct the cell shape, are disassembled and their components are used to form the mitotic spindle, which assists the correct division of chromosomes. Prometaphase is the second mitotic phase where the nuclear envelope completely breaks down. The mitotic spindle fibres elongate to specific areas on the chromosomes and the formation of a protein, kinetochore around the centromere occurs. Metaphase is the third phase where the chromosomes are aligned at the equator of the cell, attached to the mitotic spindle through kinetochores. The spindle is organised by two centrosomes at the two opposite poles of the cell. Anaphase is the fourth phase where the sister chromatids are separated into two identical ones at the centromere by the mitotic spindle. The separated chromatids are pulled towards the centrosomes at the two opposite poles by the spindle. The final phase of mitosis is called telophase where the separated chromosomes arrive at the two cell poles, and then uncoil. The nuclear membrane starts forming around each set of chromosomes. Telophase is followed by cytokinesis, a process where the cytoplasm of the parent cell divides to complete the formation of two identical daughter cells. (Figure 1.1 <sup>3</sup>) Each phase of the cell cycle contains checkpoints, where progression is halted until certain conditions are met. For instance, the checkpoints at the end of G1, S and late G2 phases allow the cell to assess DNA damage for repair or apoptosis (programmed cell death) if the damage is too extensive to repair. Understanding the cell cycle and its control may assist the development of effective therapies to treat the disease. <sup>1c</sup>



**Figure 1.1.** A cellular and molecular view of mitosis. The figure represents NIH 3T3 cells at different stages during mitosis. DNA (blue), microtubules (red) and PLK1 (green). Some representative regulators altered in human cancer are indicated. The involvement of APC/C ubiquitin ligases during different mitotic stages is indicated with thick arrows. <sup>3</sup> (Reproduced from Figure 1 Reference 3. Used by Permission of the publisher.)

*Centrosome cycle.*<sup>4</sup> The description of the cell cycle shows that centrosomes (the microtubule-organising centres of cells) play a fundamental role in regulating accurate cell division in mitosis. In mitosis, the segregation of centrosomes is linked to the segregation of chromosomes. It was noticed that there is a strong correlation between aberrant centrosome number and aberrant chromosome number in cancer cells.<sup>4b</sup>

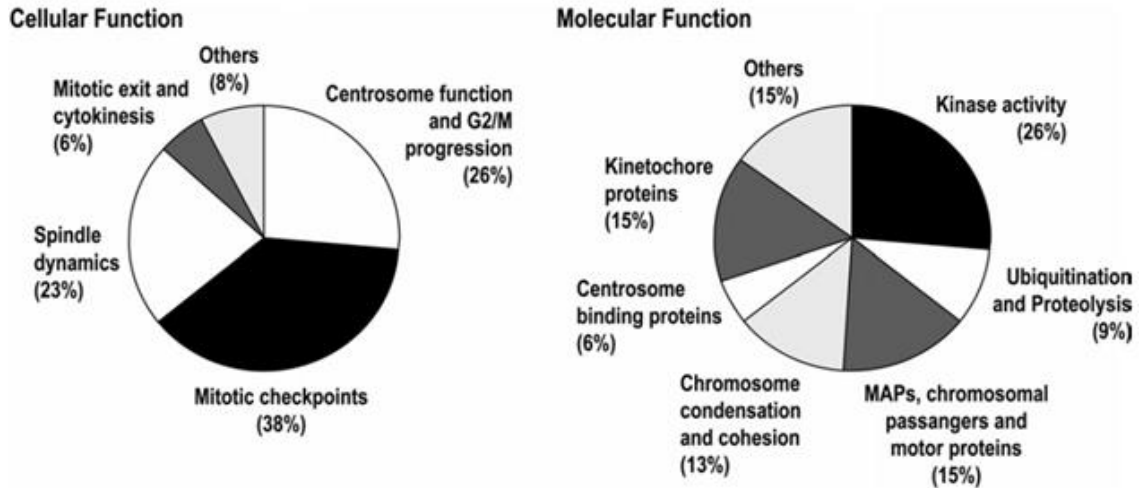


**Figure 1.2.** Centriole biogenesis. This schematic representation of the centriole duplication cycle shows centrioles (green) and pericentriolar material (PCM) (grey), with emphasis on two distinct linker structures. The G1–G2 tether (GGT; blue) connects the proximal ends of the two parental centrioles from G1 to late G2; it is important to ensure microtubule nucleation from a single microtubule organizing centre. The S–M Linker (SML; red) forms during S phase and connects the proximal end of the nascent procentriole to the lateral surface of the mother centriole. The removal of this tight connection in late M phase (disengagement) is an important element of cell cycle control of centriole duplication. Both the molecular components of the GGT and SML as well as the regulation of the formation and dissolution of these structures are expected to be distinct, although some PCM components are likely to be important for both GGT and SML. Also depicted are subdistal and distal appendages (triangles); although readily visible in electron micrographs during interphase, these appendages are difficult to visualize during M phase. In quiescent cells, the appendage-bearing centriole associates with the plasma membrane (PM) and acts as a basal body to form a primary cilium. Finally, in multi-ciliated epithelial cells, multiple centrioles form simultaneously from an amorphous structure termed the deuterosome (D).<sup>4a</sup> (Reproduced from Figure 1 Reference 4a. Used by Permission of the publisher.)

A centrosome consists of two cylindrical microtubule-based centrioles, which recruit a matrix of associated pericentriolar material (PCM). During each cell cycle, the duplication of centrioles occurs once only, beginning around the end of the G1 to S phase. The newly formed structures are called procentrioles, which elongate throughout the S and G2 phases. There are two essential types of connection between the centrioles. The progression through the centrosome cycle involves the formation and dissolution of these two connections. The first type of connection is called 'S-M linker', formed during the S phase and connecting a procentriole to its adjacent 'mother'. This linker is tight and persists until the separation of the mother-daughter centriole at metaphase. The disengagement of centriole pairs is coordinated with chromatid separation. The second type of connection is called the 'G1-G2 tether', which provides a loose connection between the proximal ends of the two parent centrioles. This spatial arrangement helps to focus the microtubule-organising activity of the centrosome. The centrosome cycle is presented in Figure 1.2 <sup>4a</sup>.

### ***Control of the cell cycle and carcinogenesis*** <sup>5</sup>

The cell cycle is controlled by many different proteins located within the cytoplasm. These regulatory proteins are products of two types of genes: proto-oncogenes and tumour suppressor genes. Genomic instability and mutation is an enabling characteristic of cancer cells. Proto-oncogenes code for various proteins, whose functions are important in the regulation of the cell cycle, signal transduction and promoting cell progression through the cycle. Proto-oncogenes can be converted into oncogenes by mutations. <sup>1a</sup> Oncogenes code for proteins that drive the cell cycle and promote continuous cell division that may lead to carcinogenesis. For example, the genetic alterations that lead to the enhanced function of the protein product or an increase in the synthesis of the protein products, which may control cell division, eventually activate the malignant transformation of cells. An important group of proteins coded by proto-oncogenes are cyclins, which are involved in the promotion of progression through the cell cycle. Cyclin-dependent kinases (CDKs) are another group of proto-oncogenic proteins. A cyclin binds with a CDK to form a complex, which phosphorylates (transfers a phosphate group to) transcription factors, proteins that bind to genes to regulate gene expression. These transcription factors activate transcription of genes required for the next stage of the cell cycle. <sup>1c</sup> Several other kinases and phosphatases have been identified as regulators of centrosomal and mitotic cycles such as Aurora A and Aurora B kinases. De Castro *et al.* in 2007 summarised mitotic alterations in cancer categorised by molecular or cellular function. (Figure 1.3 <sup>3</sup>)



**Figure 1.3.** Summary of mitotic alterations in cancer categorised by molecular or cellular function. Protein kinases (mostly centrosomal and checkpoint kinases) are significantly represented, suggesting diverse therapeutic uses in cancer. <sup>3</sup> (Reproduced from Figure 3 Reference 3. Used by Permission of the publisher.)

Tumour suppressor genes code for proteins that prevent excessive growth of a cell. In other words, these regulatory proteins inhibit cell progression through the cell cycle. There are two types of tumour suppressor genes: caretaker and gatekeeper genes. Protein products of caretaker genes e.g. BRC1 repair structurally damaged chromosomes and correct DNA mutations. Protein products of gatekeeper genes exhibit inhibitory effects on cell growth (cell cycle arrest and/or apoptosis), their function is lost in cancers. Together with oncogenes, they play major roles in carcinogenesis. The most well-known ones are p53 and the retinoblastoma (Rb) gene. In normal cells, Rb protein blocks cell proliferation through binding and inactivating a transcription factor. Ser/Thr phosphorylation of Rb protein prevents this interaction, stopping the blockage. Its inhibitory activity usually synergises with p53 protein. <sup>1b</sup> The other enabling characteristic of cancer is tumour-promoting inflammation. This may involve necrosis, a process of cell death as a result of acute injury. The necrotic cells swell and burst, releasing their contents into the local tissue microenvironment. This may signal an inflammatory response from the immune system, sending signals e.g. growth factors promoting adjacent healthy cells to divide to compensate for the loss. <sup>5,6</sup>

Hanahan and Weinberg in 2000<sup>7</sup> and in 2011<sup>5</sup> reviewed that there are six major cell alterations dictating malignant growth. One hallmark is *self-sufficiency in growth signals*, which means that cancer cells are capable of synthesising growth factors, producing their own growth signals to which they respond. Alternatively, they may signal healthy cells within the same environment to supply various growth factors. Cancer cells may also increase the number of growth receptor proteins displayed on the surface, making them hyper-responsive to growth factor ligands. Another hallmark is *insensitivity to inhibitory signals*. For example, in retinoblastoma cancer, deletion or mutation of the RB1 gene inactivates the tumour suppressor protein, Rb protein. Rb protein integrates signals from extracellular and intracellular sources to decide a cell progression or suspension in the cell cycle.<sup>8</sup> A third feature is *evasion of apoptosis*, e.g. the deletion or mutation of tumour suppressor gene p53. Another hallmark is *limitless replicative potential*. The number of divisions that a normal cell can carry out is controlled by telomeres, the two ends of each chromosome. Telomeres, containing several repetitions of the short nucleotide sequence TTAGGG, do not code for genetic information but are to protect the ends of chromosomal DNA from end-to-end fusions. They are shortened progressively over each cell cycle, ultimately determining the number of cycles a normal cell can go through. Telomerase is an enzyme that binds to telomeres, preventing the shortening process from occurring, resulting in limitless replicative potential. In cancer cells, this enzyme is reportedly expressed in significant levels. The fifth and sixth features of cancer are *sustained angiogenesis*, which supplies nutrients and oxygen to the tumour and removes metabolic waste and carbon dioxide from it, and *tissue invasion and metastasis*, which involve the physical spreading of cancer cells from the primary location to distant tissues followed by the adaptation of these cells to foreign microenvironments, resulting in colonisation and formation of secondary macroscopic tumours.

### **Principles of cancer treatment<sup>1a, b</sup>**

The development of a malignant tumour is characterised by stages of disease (tumour size and degree of spread) and by grades of disease (cellular characteristics of the cancer). Cancers are commonly categorised into stages using the **TNM** system, which is based on three factors: **t**umour size, spread to the lymph **n**odes and **m**etastasis to distant sites. In terms of grading, a low-grade tumour histologically resembles the original tissue much more than a high-grade cancer, which has gone through many changes (cellular heterogeneity).<sup>1b</sup> These two factors (stages and grades) are essential to outline an effective course of treatment for cancer.

Three traditional forms of cancer treatment are surgery, radiation and chemotherapy. Surgery and radiation are generally used in treatments of localised tumours that are confined to specific locations whereas chemotherapy is a systemic treatment. <sup>1a</sup> Surgery is primarily used to reduce or possibly remove the entire tumour. It can also be used for the purpose of preventing the risk of cancer development. Radiation therapy mostly uses high-energy X-rays to destroy tumours by causing DNA damage. Minimal damage to adjacent healthy cells is obtained by selective shielding and varying the direction of the radiation beams. Radiation therapy is based on the theory that cancer cells have reduced ability to provide protection against or repair DNA damage compared to healthy cells due to the lack of functional protein products of tumour suppressor genes. Chemotherapy is, as mentioned, a systemic treatment. It refers to the use of cytotoxic drugs to inhibit the growth of cells and kill them. Chemotherapeutic drugs can enter the bloodstream and be delivered to various sites, so chemotherapy is systemic.

Chemotherapeutic drugs often bind to molecules and interfere with their ability to carry out certain functions in particular stages of the cell cycle. Actively proliferating cells are most susceptible to chemotherapeutic drugs. Alkylating agents (e.g. cyclophosphamide) form adducts with DNA bases, disrupting DNA synthesis. Antimetabolites (e.g. methotrexate) inhibit nucleic acid synthesis. Other targets include topoisomerase, the enzyme that unwinds the DNA helix during and after replication and microtubules in mitosis. One famous example is Imatinib, a tyrosine kinase inhibitor, used in the treatment of chronic myeloid leukaemia and gastrointestinal stromal tumours. In this project, I focused on protein kinases, which are dominant in regulating cell proliferation and have been considered as a possible therapeutic target in cytotoxic chemotherapy. A background on kinases and their role in chemotherapy is presented in Section 1.2.

## **1.2. Protein kinases for cancer therapy** <sup>9,10</sup>

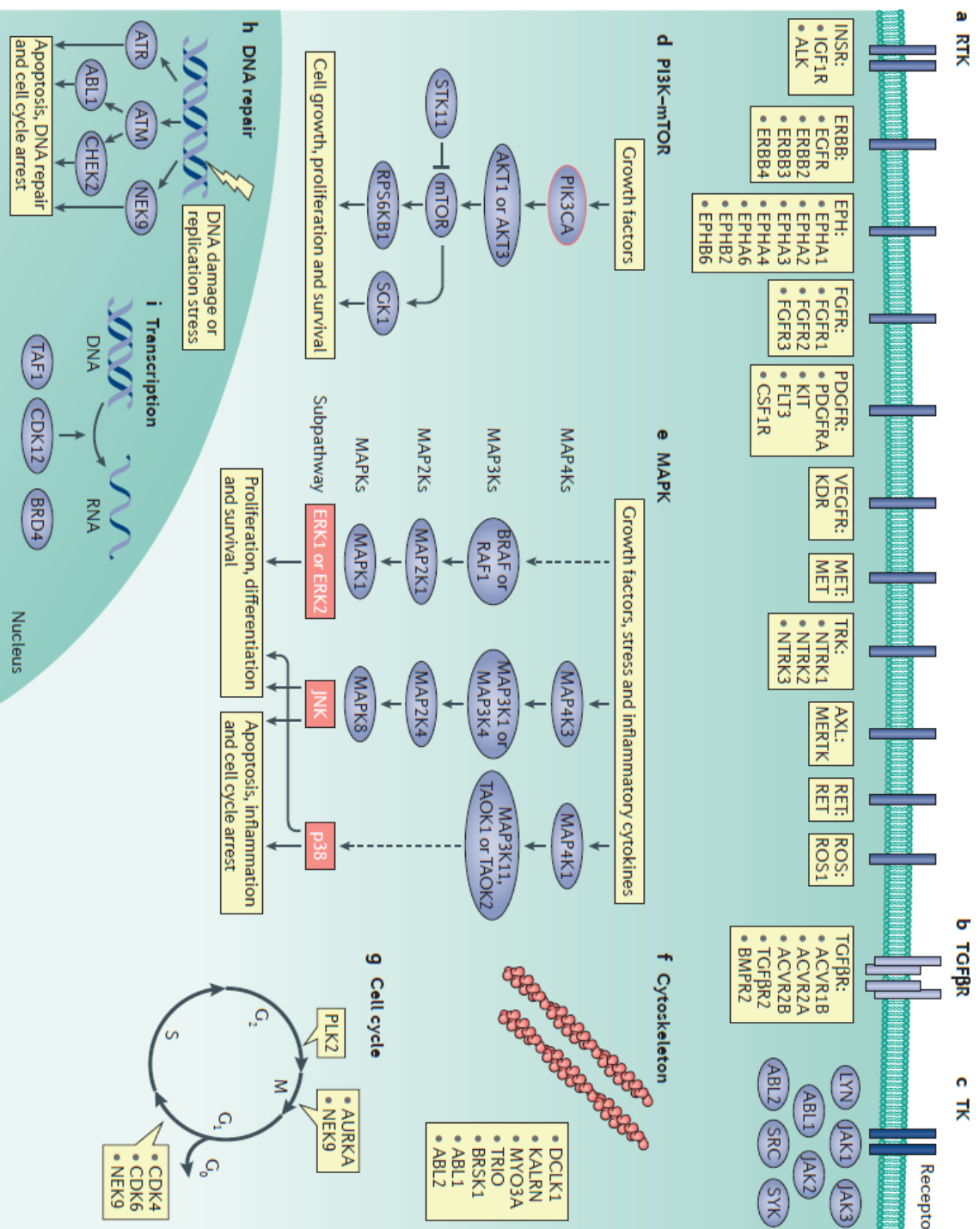
Protein phosphorylation is a common post-translational modification of proteins in signal transduction, playing an important role in cellular and extracellular processes including differentiation, membrane transport, metabolism and growth. A protein kinase is an enzyme that catalyses the transfer of the gamma phosphate group from ATP onto hydroxyl groups of various substrates such as lipids, sugars and amino acids. The reverse process is catalysed by phosphatases. In eukaryotes, the three common amino acids of proteins that are sites of phosphorylation are tyrosine, serine and threonine. Eukaryotic protein kinases are classified into three groups: tyrosine-specific protein kinases, serine/threonine-specific protein kinases and dual-specific (both tyrosine- and serine/threonine-specific) protein kinases. The majority of eukaryotic protein kinases are serine/threonine-specific protein kinases, reflecting in the ratio of cellular phosphorylation (pSer : pThr : pTyr = 1000 : 100 : 1). <sup>9,11</sup>



Protein kinases are one of the largest families of genes in eukaryotes, making up about 2% of the human genome. The classification of kinases was primarily performed by sequence comparison of their catalytic domain, supported by the information of sequence similarity and domain structure outside of the catalytic domains. The protein kinase complement of the human genome, the human kinome, contains 518 eukaryotic protein kinase genes according to the latest count.<sup>12</sup> These are subdivided into seven families of typical and seven families of atypical kinases. The typical kinases are those that contain a conserved catalytic domain belonging to the eukaryotic protein kinase superfamily. Atypical kinases lack sequence similarity to the typical kinase catalytic domain but have been shown experimentally to exhibit protein kinase activity. In general, members of the same kinase family have the same domain structure. The seven families of the typical kinases are AGC (PKA, PKG and PKC) kinases, CAMK (CA<sup>l</sup>cium/Cal<sup>M</sup>odulin-dependent) kinases, CMGC (the C<sup>Y</sup>clin-dependent kinases, M<sup>A</sup>P kinases, G<sup>L</sup>ycogen synthase kinases, C<sup>A</sup>sein kinases 2) kinases, CK1 (C<sup>A</sup>sein K<sup>I</sup>nase 1) kinases, STE kinases, TKL (the T<sup>Y</sup>rosine K<sup>I</sup>nase L<sup>I</sup>ke) kinases and TK (T<sup>Y</sup>rosine K<sup>I</sup>nase) kinases. Apart from the TK family, the other 6 families are serine/threonine-specific protein kinases. About 10% of all eukaryotic kinase domains (50 domains) lack catalytic activity due to sequence change in essential motifs and are called pseudo-kinases. However, it was postulated that they retained other non-catalytic functions.<sup>13</sup> Some may act as modulators of other catalytic domains. For example, GCN2 (G<sup>E</sup>neral C<sup>O</sup>ntr<sup>O</sup>l N<sup>O</sup>n-derepressible 2), a serine/threonine-specific protein kinase that senses amino acid deficiency through binding to uncharged transfer RNA and plays a key role in modulating amino acid metabolism as a response to nutrient deprivation, has dual domains; one of which is inactive and may regulate the active domain.<sup>14</sup>

Studies in cancer genomics have demonstrated the importance of protein kinases in cancer development. Early work by Futreal *et al.* in 2004<sup>15</sup> on the Cancer Gene Census, a literature based census of genes that were mutated and causally implicated in cancer development showed 27 of 291 “cancer genes” encoded protein kinases. In 2010, Kan *et al.*<sup>16</sup> studied 441 specimens derived from various human cancers and revealed that mutations were found in 157 of the 230 investigated protein kinases. Recent studies of large-scale cancer genome sequencing, including The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC), to address somatic genetic alterations characteristic of common human cancer types revealed that certain protein kinases were mutated at relatively high frequency (>10%) in one or more human cancers.<sup>17</sup> Some notable examples include BRAF in thyroid cancer, skin cutaneous melanoma (60% and 51% respectively) and Fms-related tyrosine kinase 3 in acute myeloid leukemia (27%). These studies were also able to detect cancer drivers (a term which refers to both oncogenes and tumour suppressors) that mutated at intermediate (1-10%) or low (<1%) frequency. Fleuren and Zhang *et al.*<sup>10</sup> in 2016 summarised the functional annotation of the identified 91 kinase drivers in cancer, including signalling pathways via transmembrane

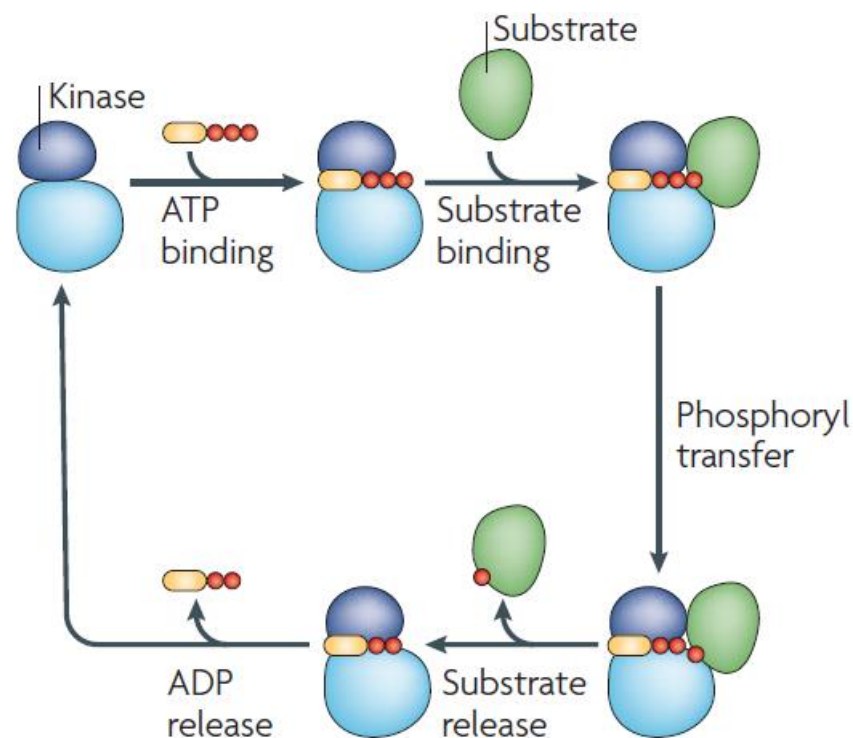
receptors, signalling pathways via cytoplasmic proteins, signalling to the nucleus, cell-cycle progression, DNA repair and regulation of apoptosis. (Figure 1.4) Receptor tyrosine kinases are major kinase drivers with about half of the 58 known tyrosine kinases being among the identified kinase drivers. (Figure 1.4a) These transmembrane proteins are single subunit receptors, composed of an extracellular N-terminal region and an intracellular C-terminal region. The intracellular region contains a highly conserved catalytic site, responsible for the activation of the kinase through the phosphorylation process. The extracellular region contains a ligand-binding site to bind extracellular ligands such as hormones and growth factors (e.g. epidermal growth factors - EGF, vascular endothelial growth factors - VEGFR and platelet-derived growth factor - PDGF). EGF takes part in the regulation of cell growth and differentiation. PDGF regulates cell growth and development. VEGFR is involved in the creation of blood vessels. As a result, these receptor tyrosine kinases regulate various cellular responses such as cell proliferation and migration. Another type of kinase drivers reported was the serine/threonine-specific kinases that were parts of the receptor complexes for transforming growth factor- $\beta$  (TGF $\beta$ R) listed in Figure 1.4b. The non-receptor tyrosine kinases, which transduce and/or regulate signals downstream of specific cell surface receptors, were also listed as cancer drivers. (Figure 1.4c) These include members of the ABL, Src, and JAK kinases. ABL1 was also involved in DNA repair process and in the regulation of the organisation of the cytoskeleton. (Figure 1.4h and 1.4f) Six kinase drivers involved in the PI3K/AKT/mTOR signalling pathway in the cytoplasm, which promotes cell proliferation, survival and protein synthesis are listed in Figure 1.4d. Other key signalling pathways include MAPK cascades, which transduce signals from the cell surface such as growth signals, stress and inflammatory signals to the DNA and are essential in the regulation of normal cell proliferation, survival and differentiation. As a result, it is understandable that many members of the MAPK kinase family were also listed as cancer drivers (Figure 1.4e) Other kinase drivers include the regulators of the cell cycle such as cyclin-dependent kinases CDK4 and CDK6, which promote cell-cycle progression and AURKA and NEK9, which are known to regulate spindle assembly and chromosome separation and segregation.



**Figure 1.4.** Assignment of driver kinases to core cellular pathways and processes. Names identified in blue ovals indicate particular driver kinases. Dashed lines indicate that the specific kinases involved at these steps in the pathway have not been identified as drivers. Assignment of driver receptor tyrosine kinases (RTKs) to particular RTK families (4a). Driver kinases that act as receptors for specific transforming growth factor- $\beta$  (TGF $\beta$ ) superfamily members (TGF $\beta$ Rs) (4b). Non-receptor tyrosine kinases (TKs) that act as drivers, highlighting their role downstream of particular types of cell surface receptor (4c). Particular driver kinases are also associated with the PI3K-mTOR (4d) and MAPK (4e) pathways, regulate cytoskeletal organization (4f) or the cell cycle (4g), or function in DNA repair responses (4h) or gene transcription (4i). <sup>10</sup> (Reproduced from Figure 2 Reference 10 Used by

*The structure and catalytic mechanism of eukaryotic protein kinases*<sup>11</sup>

All the protein kinases share a common protein folding and catalytic mechanism. The catalytic domains of protein kinases are usually about 250 amino acids long and highly conserved, containing a  $\beta$ -stranded N lobe and an  $\alpha$ -helical C lobe connected together by a short hinge region. The hinge region is where the protein kinases bind ATP. In other words, the ATP binding site of a protein kinase is positioned in a cleft between the N- and C-terminal of the protein. The catalytic domain is defined mainly by the hinge region, where ADP/ATP binds and the activation loop, which is highly disordered. The hinge region is conserved among kinases. The adenine group of ATP binds to the hinge region through hydrogen bonding (H-bond) with the backbone  $-C=O$  and  $-NH$  groups of the hinge region's residues pointing into the cleft. The ribose and triphosphate group of ATP bind to a hydrophilic channel, extending to the activation loop.<sup>18</sup> The protein substrate binds along the cleft and a set of conserved residues within the kinase catalytic domain catalyse the transfer of the terminal  $\gamma$ -phosphate of ATP to the hydroxyl of the Ser, Thr or Tyr residue of the substrate. The summary of this process was illustrated in a review of the mechanisms of specificity in protein phosphorylation by Ubersax and Ferrell. (Figure 1.5<sup>11</sup>)

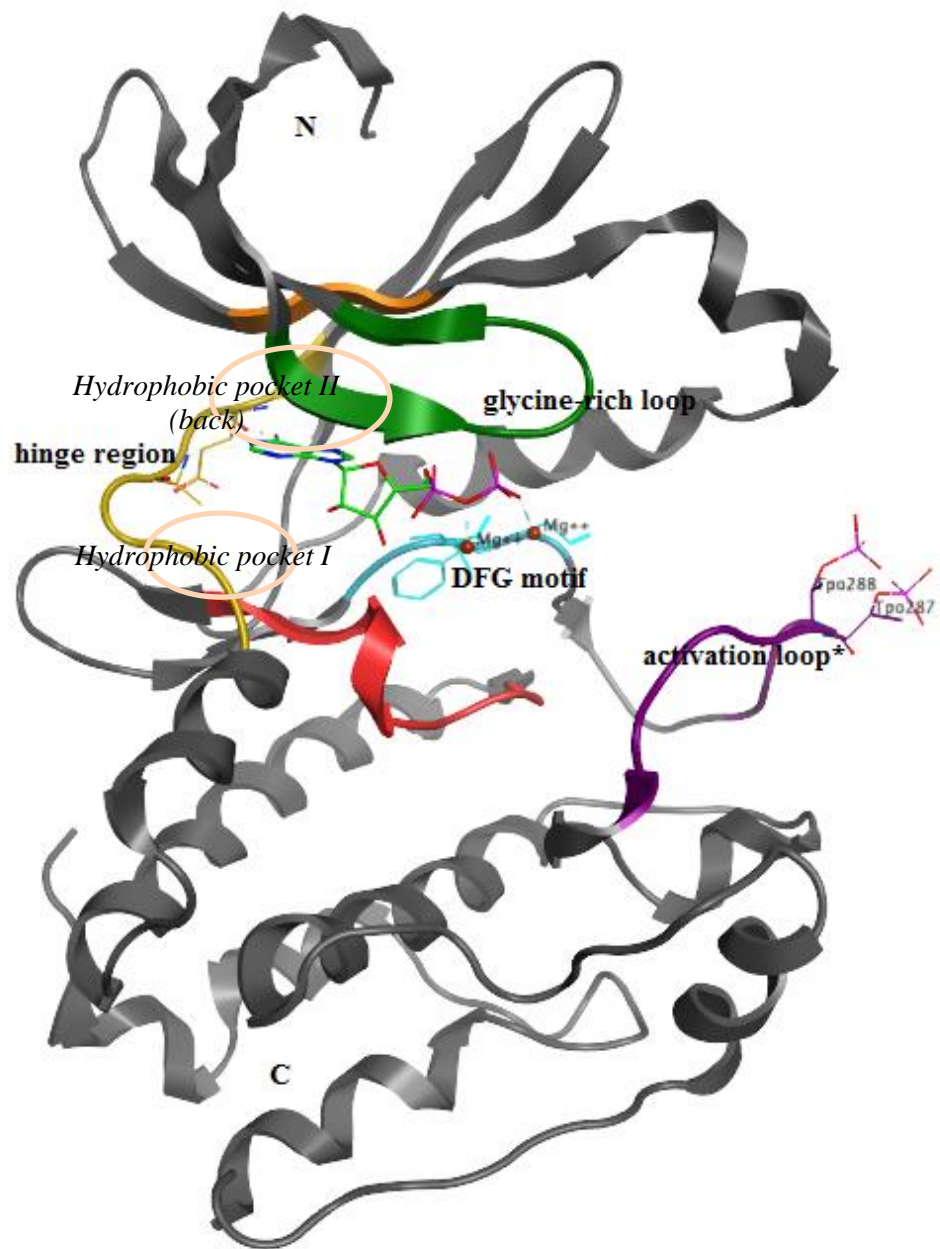


**Figure 1.5.** The basic catalytic cycle for substrate phosphorylation by a protein kinase.<sup>11</sup>

(Reproduced from Figure 1a Reference 11. Used by Permission of the publisher.)

The activation loop plays an essential role in the regulation of enzyme activity by providing the platform for binding and positioning of the hydroxyl group residue of the peptide substrate. It generally starts with a conserved three amino acid DFG motif. This region is extremely flexible

and can assume many different conformations as it regulates the on and off state of the kinase. As such it is very difficult to capture one single conformation of the activation loop in the crystal structures and these therefore appear as a disordered region. In the inactive state, the activation loop usually points upwards and towards the cleft, partly blocking the catalytic domain. Phosphorylation of one or more residues in this region generally pulls the activation loop downward, and the Phe residue of the DFG motif moves from the DFG-out into the DFG-in position (i.e. being buried in a hydrophobic pocket in the groove between the two lobes of the kinase). This starts the activation of the enzyme. Figure 1.6 illustrates an example of the ATP binding site of Aurora A kinase (AURKA) where a nearly complete conformation of the activation loop is captured (PDB: 1OL7).<sup>19</sup> In this structure, ATP molecule is bound to the hinge region of AURKA by hydrogen bonding between the adenosine moiety and the backbone  $-C=O$  and  $-N-H$  of A211 and E213. Two residues located in the activation loop, T287 and T288 were phosphorylated. This is the most complete crystal structure found among the reported crystal structures of AURKA kinase in RCSB PDB. However, even in this structure, a few side-chains of the residues composing the activation loop were missing due to its flexibility.



**Figure 1.6.** ATP binding site of a protein kinase. Example: human AURKA kinase, PDB: 1OL7. Molecule: ADP (C atoms in green), DFG in. Hinge region (yellow), glycine-rich loop (green), DFG motif (cyan), activation loop\* (purple): phosphorylated threonine residues.

The depth and hydrophobicity characteristics of kinases often increase the binding preference for substrates. These features help maintain the specificity of kinases in signalling pathways. For example, tyrosine kinases generally have a deeper catalytic cleft than serine/threonine kinases to accommodate the sidechain of the tyrosine residue pointing into the catalytic cleft. The same feature was found in tyrosine phosphatases and serine/threonine phosphatases.<sup>20</sup> Outside of the catalytic domain, some protein kinases contain allosteric sites which may positively or negatively contribute to the regulation of kinase activity. An example of this is the

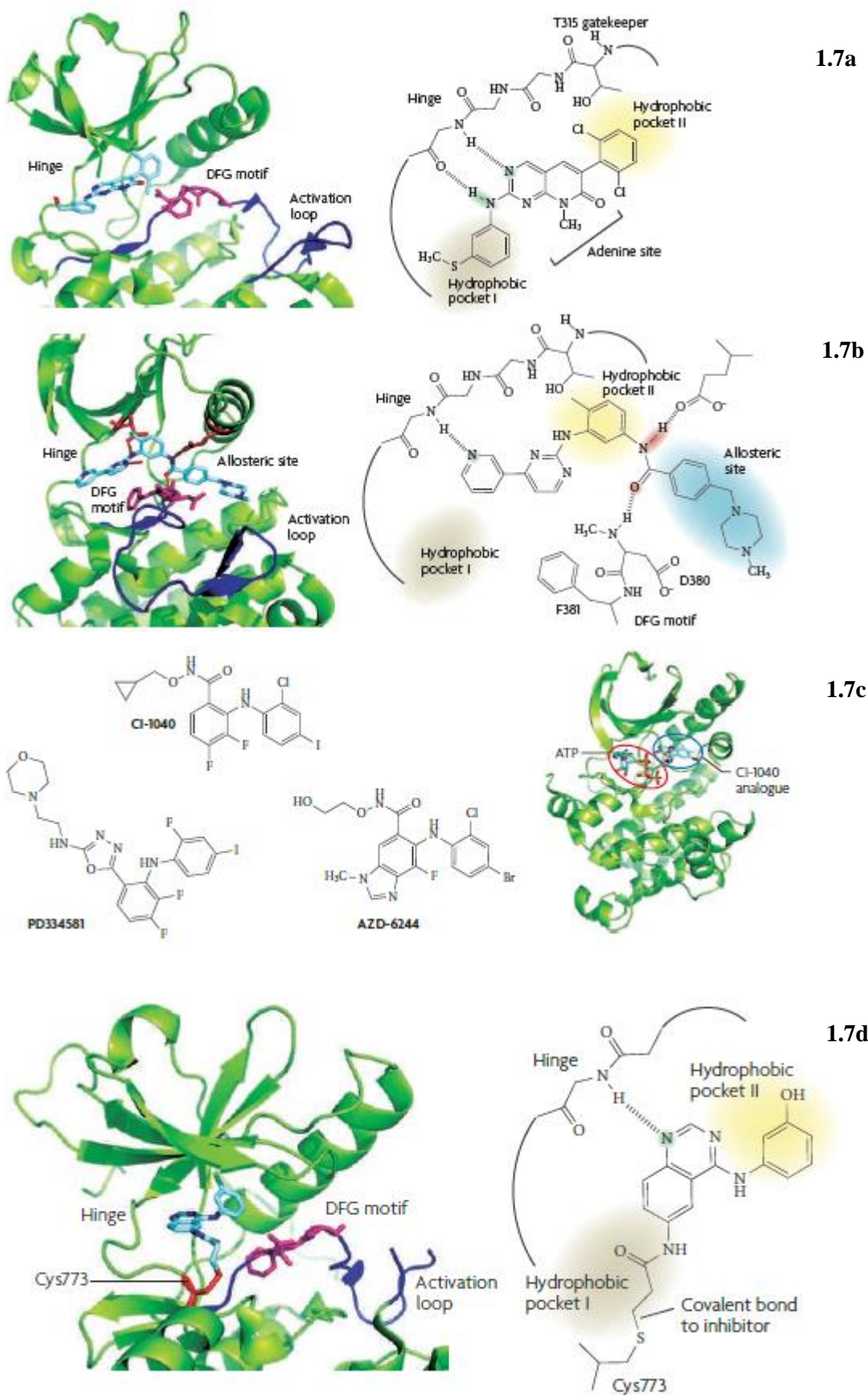
PDK1-interacting fragment (PIF) pocket of PDK1 kinase of AGC family, which recruits downstream substrate kinases to allosterically regulate PDK1 activity.<sup>21</sup>

#### *Kinase inhibitors in chemotherapy*<sup>9, 18, 22</sup>

As many kinases have been found to take part in the process leading to tumour cell proliferation and survival, oncology drug discovery has been exploiting them as therapeutic targets for cancer therapy. The current kinase inhibitors can bind either covalently or reversibly (non-covalent) to kinases. They are generally categorised into 4 types according to different modes of action elucidated by X-ray crystallography.<sup>9, 18</sup>

*Type 1 inhibitors, ATP-competitive inhibitors* bind to the active conformation of the kinase (DFG-in conformation). They typically consist of a heterocyclic system that occupies the hinge region and the extended sidechains that occupy hydrophobic pockets nearby. An example of this is inhibitor PD166326 of ABL1. (Figure 1.7a) *Type 2 inhibitors*, such as Imatinib and Sorafenib, still have contact with the hinge region but stabilise the inactive conformation of the kinase (DFG-out conformation). (Figure 1.7b) *Type 3 inhibitors, allosteric inhibitors*, as the name suggests, bind to an allosteric site, distant from the hinge region but adjacent to the ATP binding site, and they may differ from kinase to kinase.<sup>23</sup> As a result, these may exhibit a higher degree of selectivity by exploiting binding sites and regulatory mechanisms that are unique to a particular kinase. An example of this is CI-1040, inhibiting MEK1 and MEK2. (Figure 1.7c) *Type 4 inhibitors, covalent or irreversible inhibitors*, may form a covalent bond with a residue in the active site. The common target kinase is one with a reactive cysteine residue within the ATP binding site. Due to the nature of strong covalent bonding, these inhibitors may exhibit higher potency and improved inhibition kinetics than those of other types. There are a number of marketed drugs which covalently bound to their targets.<sup>24</sup> However, there is also a concern of their potential for off-target reactivity which may lead to potential toxicity. This explains the reluctance to apply a covalent mode of action in drug discovery programmes. For covalent inhibitors, the target sites are usually the Lys or Cys residues located in or around the ATP binding site. An example of this type is inhibitor 34-JAB against EGFR kinase (PDB: 2J5F). (Figure 1.7d)





**Figure 1.7.** Types of kinase inhibitors. Kinase inhibitor-Protein interaction is depicted by ribbon structure and chemical structure. a) ABL1 in complex with the type 1 ATP-competitive inhibitor PD166326 (PDB: 1OPK); b) The DFG-out conformation of the activation loop of



ABL1 (dark blue) with the type 2 inhibitor Imatinib (PDB: 1IEP). The allosteric site exposed in the DFG-out conformation is indicated in blue shaded area; c) Allosteric kinase inhibitors. CI-1040, PD334581 and AZD-6244 of MEK1 kinase. CI-1040 binds to MEK1 (PDB: 1S9I); d) The DFG-in conformation of EGFR with covalent inhibitor 34-JAB (PDB: 2J5F). The covalent bond is indicated between C773 and the inhibitor. <sup>18</sup> (Modified version of Figure 1, Figure 3 and Supplementary information S2 in Reference 18. Used by Permission of the publisher.)

Of the 91 identified kinase drivers, 20 kinases (22%) are approved by the FDA (US Food and Drug Administration) as drugs. Almost all of them are tyrosine kinases. Twenty-three of the kinases (25%), which are in clinical trials, are dominated by tyrosine kinases but also consist of serine/threonine-specific kinases. The most well-known example of a tyrosine kinase inhibitor as a successful anticancer agent, and also the first marketed kinase inhibitor (in 2001), is Imatinib/Gleevec. It is a potent inhibitor of ABL kinases, including the BCR-ABL fusion protein uniquely expressed in leukemia cells. It also inhibits two other receptor tyrosine kinases, c-Kit and PDGFR via binding to the inactive conformations of these kinases. Sunitinib, an FDA-approved marketed drug by Pfizer is another receptor tyrosine kinase inhibitor. It is used for the treatment of renal cell carcinoma and imatinib-resistant gastrointestinal stromal tumors. Another example of the marketed tyrosine kinase inhibitors as anticancer agents is Lapatinib, which dually inhibits EGFR and HER2/neu (Human EGFR type 2). <sup>25</sup> One common biochemical characteristic of the three drugs is that these inhibitors all compete with ATP for the binding to the catalytic site of the receptor tyrosine kinases. In fact, there are still few inhibitors that interact with the region, above which the ribose of ATP resides, and/or interact with the activation loop. It is open for investigation as to whether there is any benefit in terms of inhibitory effect and selectivity from interactions with these regions. For the development of the next generation of kinase inhibitors, it is desirable to achieve a more selective kinase inhibitor and to have a better understanding of the structure of the extended ATP binding site (activation loop), apart from the hinge region.

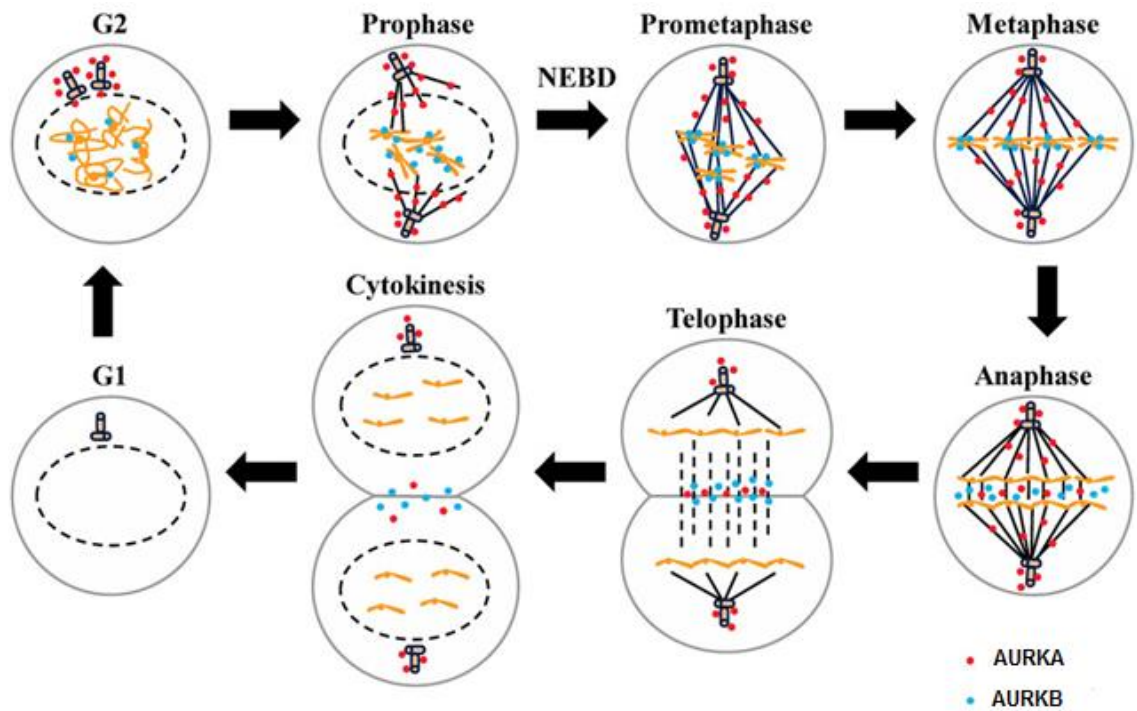
Besides the dominance of the tyrosine kinase inhibitors in clinical trials, inhibitors of the serine/threonine-specific kinases that are members of the AGC family have also presented a significant portion in clinical trials. One example are the Aurora kinases, which have been mentioned to play significant roles in regulating cell-cycle progression. The first strong supporting evidence for this was the reported preclinical results for VX680 compound in 2004, which demonstrated that Aurora kinase inhibitors could prevent tumour progression in human models of solid tumour cancers. <sup>26</sup> This project focuses on Aurora A kinase as a therapeutic target for cancer.

### 1.3. Aurora A kinase (AURKA) for cancer therapy <sup>27</sup>

The Aurora kinase family consists of three members: Aurora A (AURKA), Aurora B (AURKB) and Aurora C (AURKC). The three are serine/threonine-specific kinases and share a highly conserved catalytic domain containing auto-phosphorylating sites. AURKA and AURKB are commonly expressed in most cell types whereas AURKC is found in the testes. Functionally, AURKC plays a role in spermatogenesis. The three Aurora kinases also function as mitotic kinases. AURKA and AURKB take part in the regulation of the cell-cycle progression from G2 to cytokinesis. Consequently, overexpression or amplification of AURKA and AURKB has been detected in many types of cancer such as breast, colorectal, pancreatic, ovarian, oesophageal, gastric and bladder cancers. <sup>28</sup> AURKC is also found as a chromosomal passenger protein (like AURKB) during mitosis. However, there is little information regarding its precise function and its role in carcinogenesis. Structurally, AURKA kinase has been extensively studied compared to AURKB and AURKC. There have been a number of crystal structures reported for human AURKA. However, there is currently no human AURKB crystal structure reported. <sup>29</sup> Hence, this project will focus on interfering with AURKA activity as a therapeutic target for cancer treatment.

#### *Localisation and Functions of AURKA* <sup>30, 31</sup>

In the cell cycle, AURKA is first detected at the centrosomes during the S phase. The amount of AURKA rises at the centrosomes during late G2 phase and the kinase becomes activated. The amount of active AURKA reaches its peak and is detected on the bipolar spindles and spindle poles during the prometaphase and metaphase. After that, the majority of AURKA becomes inactivated and degraded in anaphase and only a small amount is found on the centrosomes and the spindles particularly the spindle midzone and remains there through telophase and cytokinesis. A brief illustration of the expression patterns and localisation of AURKA and AURKB was presented in Yan *et al.* review in 2016. (Figure 1.8) <sup>27</sup>



**Figure 1.8.** The expression levels and localisation of AURKA and AURKB in the cell cycle. In the G1 phase, AURKA is rarely detectable. During S phase, a small proportion of AURKA is first detected at centrosomes. During late G2 phase, AURKA accumulates evidently at the centrosomes and becomes activated. During prometaphase and metaphase, active AURKA localizes to bipolar spindles and spindle poles following nuclear-envelope breakdown (NEBD). At the metaphase-anaphase transition, the majority of AURKA is inactivated and degraded. A small fraction of AURKA remains on the centrosomes and the spindles at the onset of anaphase, and localises to the spindle midzone and centrosomes during late anaphase and telophase/cytokinesis. In mammalian cells, AURKB is first found on pericentromeric heterochromatin during late S phase, and maintains its activity throughout mitosis with its protein level peaking in the G2/M phase. During prophase, AURKB is targeted to the heterochromatin and is further enriched at the inner centromeres during prometaphase before the metaphase-to-anaphase transition. At the onset of anaphase, AURKB relocates to spindle microtubules and then to the equatorial cell cortex during cytokinesis.<sup>27</sup> (Modified version of Figure 1c in Reference 27. Used by Permission of the publisher.)

The review of AURKA by Yan *et al.* in 2016<sup>27</sup> also summarised three key functions of AURKA in mitosis, asymmetric division and cilia dynamics. In mitosis, AURKA has been known to play a role in the regulation of mitosis entry, chromosome alignment, centrosome separation and maturation and of spindle assembly.<sup>27, 30</sup> AURKA participates in the G2-M checkpoint and the regulation of mitosis entry. Prior to the M phase, AURKA, coupled with hBora protein, is involved in the phosphorylation and activation of polo-like kinase 1 (PLK1), which specifically participates in mitotic entry and mitotic progression by triggering the

activation of CDK1/Cyclin B complex, a mitosis-promoting factor. PLK1 is also indicated in centrosome maturation. It helps targeting AURKA to the centrosomes.<sup>32</sup> AURKA promotes centrosomal microtubule stabilisation by inducing the phosphorylation of transforming acidic coiled-coil-containing protein (TACC), which leads to its complex formation with XMAP215.<sup>33</sup> AURKA also phosphorylates Ajuba, which in turn further activates AURKA at late G2, forming a positive feedback loop to promote mitosis entry. In prophase, AURKA induces the phosphorylation of Eg5, a kinesin-like motor, to enhance centrosome separation and spindle assembly. Another substrate of AURKA is TPX2, upon binding to which AURKA adopts an active conformation. This activation is required in localisation of the kinase to spindle microtubules but not to spindle poles.<sup>34</sup> In metaphase, AURKA is involved in regulating chromosome alignment and kinetochore function by phosphorylating centrosome protein A, CENP-A, which triggers the AURKB-dependent phosphorylation of CENP-A and kinetochore function and CENP-E, which leads to the delivery of protein phosphatase 1 (PP1) to the kinetochore to stabilise the biorientation of chromosomes.<sup>35</sup> At the spindle checkpoint, AURKA is involved in promoting the activation of APC/Cdc20 complex, which assists cell-cycle progression. At the end of mitosis, degradation of AURKA by the APC/Cdh1 complex ensures proper cytokinesis and mitotic exit.<sup>36</sup>

Besides its important roles as a mitotic kinase, AURKA is also indicated in the regulation of primary cilia. The cilia are microtubule-based structures that protrude from the cell surface, and function as sensors for mechanical and chemical environmental cues that regulate cellular differentiation or division. AURKA negatively regulates ciliary dynamics in proliferating cells, promoting the disassembly of the cilia at the re-entry of the cell cycle and continuously suppresses the regeneration of the cilia during cell proliferation. These are achieved by the interaction of AURKA with the enhancer of filamentation 1 (HEF1/NEDD9).<sup>37</sup>

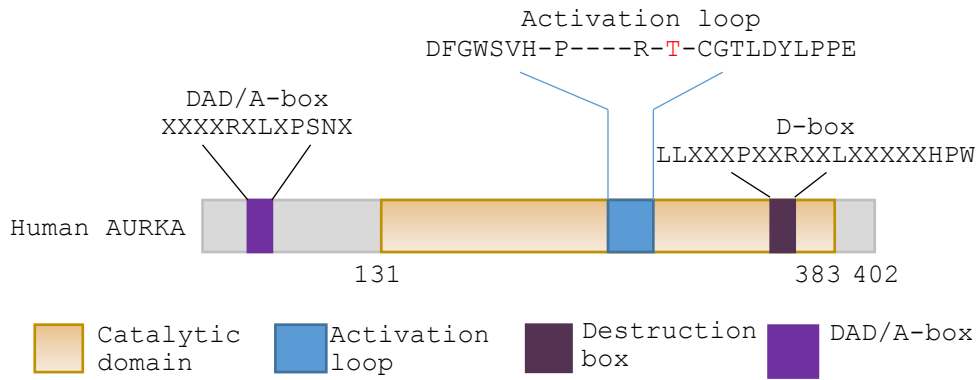
Studies on *Drosophila* neuroblasts suggested the involvement of AURKA in the regulation of cell polarity and asymmetric division. Asymmetric cell division is a process by which a cell divides to generate two daughter cells that are already different at birth. Such divisions are widely used by invertebrates to generate cellular diversity during development. The astonishing cellular diversity in the central nervous system arises from neural progenitors, which can undergo different modes of symmetric and asymmetric divisions to self-renew as well as produce differentiated neuronal and glial progeny. *Drosophila* CNS neural progenitor cells, neuroblasts, have been utilised as a model to stimulate the understanding of the processes of asymmetric division, generation of neuronal lineages and, more recently, stem cell biology in vertebrates.<sup>38</sup> Wirtz-Peitz *et al.* showed that AURKA negatively regulated Notch signalling, which was an important cell-cell communicating pathway and participated in neuronal function and differentiation.<sup>39</sup> Loss of AURKA might result in the suppression of self-renewal of neuroblasts and the promotion of neuronal differentiation.<sup>40</sup>

### *AURKA in cancer cells*

The aberrant expression and localisation of AURKA was shown to strongly link to tumorigenesis. In cancer cells, AURKA was reported to be overexpressed and distributed in both the nucleus and cytoplasm, regardless of their cell-cycle phases.<sup>41</sup> One oncogenic role of AURKA is its participation in the uncontrolled proliferation of many cancer cells. The inhibition of AURKA in ovarian cancer cells was reported to promote the expression of the tumour suppressor pRb, attenuating the G1/S transition.<sup>42</sup> The interaction between AURKA and tumour suppressor p53 was found to relate to uncontrolled cell growth. AURKA inhibitors could activate p53, resulting in proliferation arrest and senescence in melanoma tumour cells. Besides, other oncogenes may stimulate uncontrolled proliferation by inducing AURKA. For example, BCR-ABL promotes chronic myeloid leukaemia cell proliferation by activating AURKA and AURKB.<sup>43</sup> Overexpression of AURKA may lead to aneuploidy and genomic instability by disrupting the cell-cycle checkpoints. Studies showed that overexpressed AURKA induced the activation of CDK1, negating the G2/M DNA damage checkpoint.<sup>44</sup> Another oncogenic role of AURKA is supporting cancer cell survival and chemoresistance through activating the PI3K/AKT/GSK3 signalling cascade. For example, in experimental colon carcinogenesis, inhibition of AURKA-induced PI3K/AKT/GSK3 signalling cascades led to apoptosis.<sup>45</sup> In addition, AURKA promotes tumour cell migration and metastasis by activating several oncogenic signalling pathways such as AKT<sup>46</sup> and mitotic centromere-associated kinesis (MCAK) pathways<sup>47</sup>.

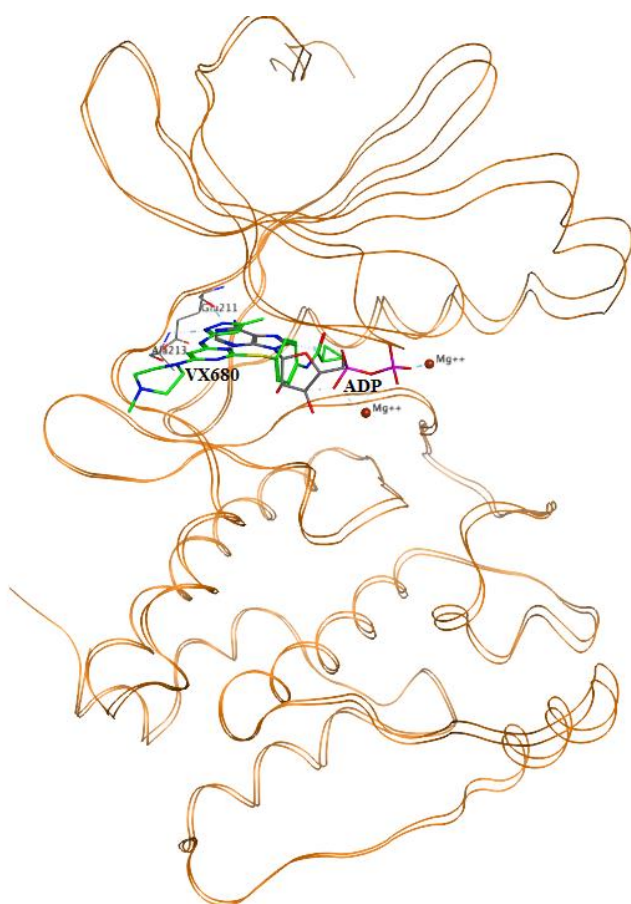
### *Structure of AURKA*<sup>30</sup>

AURKA lies within a region of chromosome 20q13 and AURKB is located at chromosome 17p13.1<sup>48</sup> The C-terminal region of AURKA contains a short amino acid motif called the destruction box, D-box. The D-box is recognised by the anaphase-promoting complex/cyclosome (APC/C) complex for degradation. It is only functional in the presence of the D-box-activating-domain, DAD/A-box. Phosphorylation of the DAD/A-box results in the resistance of AURKA to APC/C-induced degradation. (Figure 1.9)

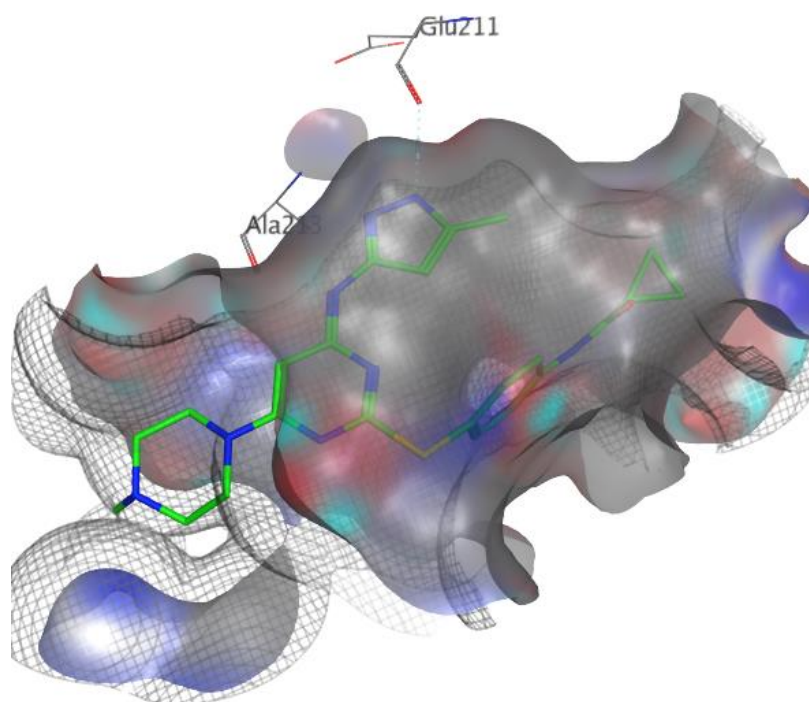


**Figure 1.9.** Structure and domains of human AURKA. <sup>30</sup>

The catalytic domain of the Aurora kinases is highly conserved. In other words, the catalytic domain of AURKA shares several characteristics with those of AURKB and AURKC. Following the general folding of kinases, the catalytic domain of Aurora kinases consists of the conserved *C*-terminal domain and short *N*-terminal domain, linked together by a hinge region which forms a part of the active site. AURKA displays high flexibility during activation, particularly in the activation loop of the catalytic domain. Its structure was presented in Figure 1.10. In addition, using ATP-competitive inhibitors such as VX680 (tozasertib) scaffold, the ATP binding site around the hinge region is explored. (Figure 1.10, PDB: 3E5A) <sup>49</sup>

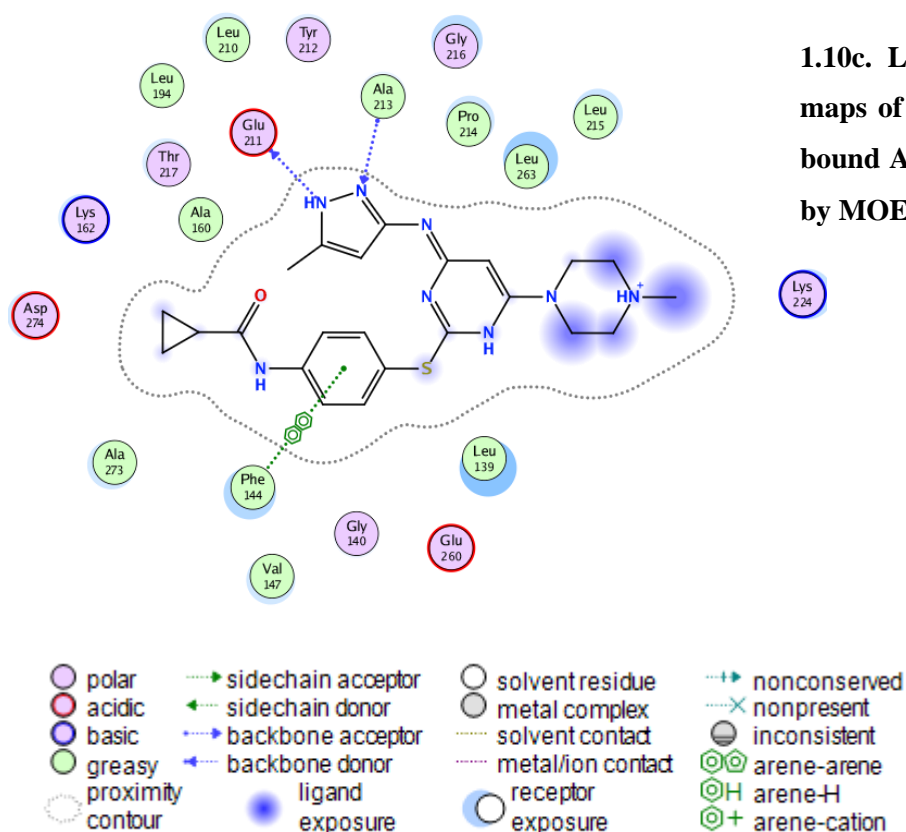


**1.10a. VX680 in phosphorylated TPX2-bound AURKA (PDB: 3E5A), superimposed with ADP/TPX2 AURKA (PDB: 1OL5). VX680 (C atoms in green), ADP (C atoms in grey)**



**1.10b. Surface interaction maps of VX680 in TPX2-bound AURKA generated by MOE. VX680 (C atoms in green). Hydrophobic area (grey), polar areas (red and cyan). Calculated van der Waals interactions (mesh).**

**1.10c. Ligand interaction maps of VX680 in TPX2-bound AURKA generated by MOE.**



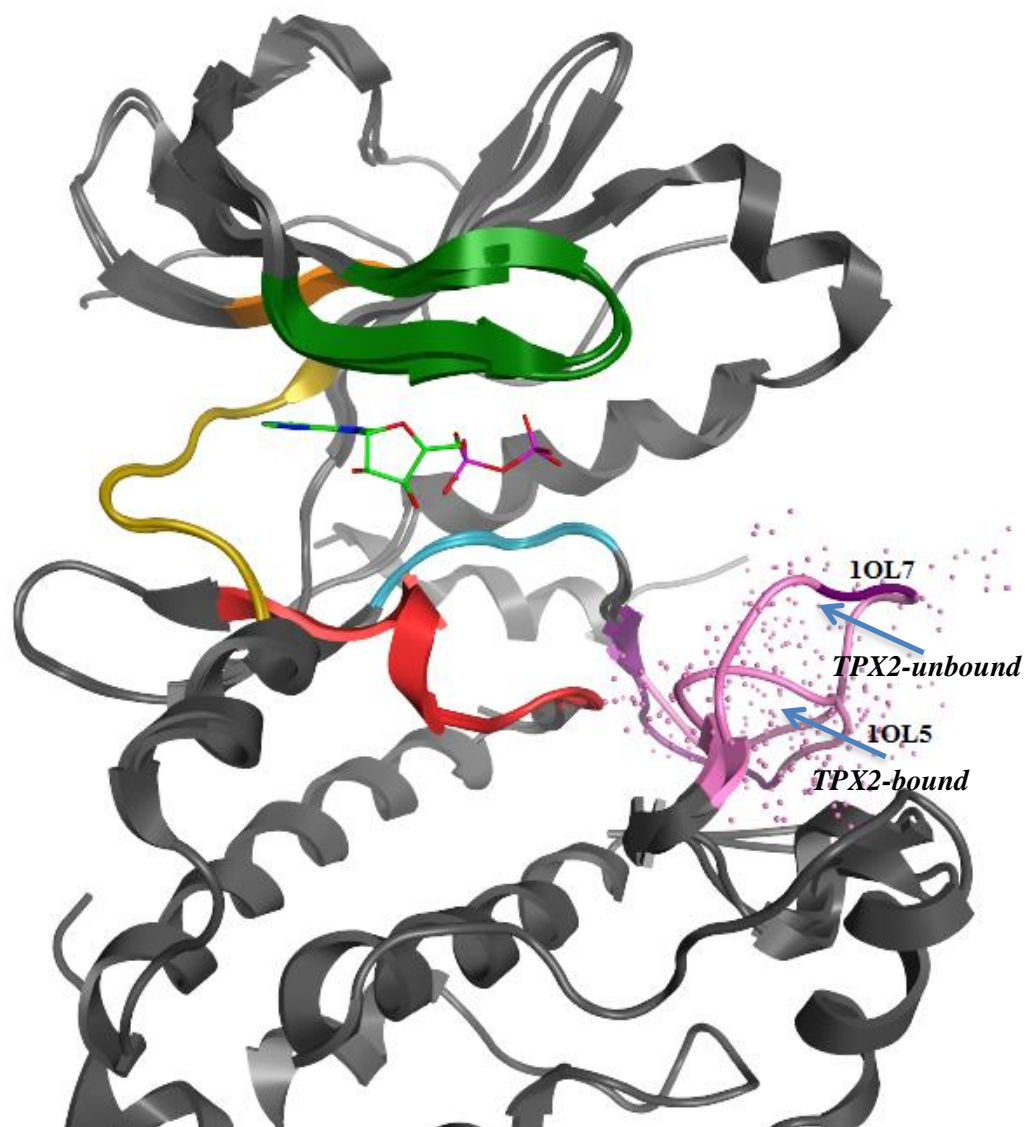
**Figure 1.10.** Crystal structure of VX680 within the ATP binding pocket of AURKA. (PDB: 3E5A)

In detail, the catalytic domain consists of 5 regions: a kinase hinge region where the main heteroaromatic core (adenine) sits, a ribose-binding region, a buried region (back pocket) for nonpolar interactions, a solvent accessible region (where the piperazine group of VX680 points) and a phosphate binding region that extends to the activation loop. (Figure 1.6) E211 and A213 of the hinge region form direct hydrogen bonds with the main scaffold of the inhibitors. The buried region usually forms hydrophobic interactions, whereas the solvent accessible region may form polar interactions and fixes the inhibitors in a particular conformation. In a biological environment, the phosphate group exists as a negatively charged species so the group designed to reach the phosphate binding region may form ionic interactions with the residues. The sizes of the chemical groups, which form the inhibitor scaffold are also important. For example, the buried region can only accommodate small groups, whilst an extended chain is probably required for better interactions with the phosphate binding region. The activation loop domains of AURKA, AURKB and AURKC are less conserved than the hinge region where ATP is bound. The conserved consensus sequence is DFGWSxxxxxxxxRxT<sup>288</sup>xCGTxDYLPPE.<sup>50</sup> Phosphorylation of this sequence at T288 usually results in activation of the kinases.



### *The regulations of AURKA activities*

The activities of AURKA are regulated by phosphorylation and dephosphorylation. Studies on *Xenopus* oocyte maturation revealed there are three main phosphorylation sites of AURKA in metaphase S53, T295, and S349, equivalent to S51, T288 and S342 in human AURKA.<sup>31, 51</sup> Mutation of S53 to Asp results in *in vitro* inhibition of AURKA destruction.<sup>52</sup> Activity of AURKA is also controlled by other proteins present at the centrosomes and spindles. In particular, AURKA is activated by TPX2, a target protein for *Xenopus* kinesin-like protein 2, which is a microtubule-associated protein required for spindle stability and AURKA localisation.<sup>19</sup> TPX2 was demonstrated to bind and stimulate auto-phosphorylation of AURKA T295.<sup>53, 54</sup> Interactions of TPX2 with NuMA, nuclear mitotic apparatus protein, and Ran (GTPase) regulate mitotic spindle assembly.<sup>19</sup> The phosphorylated activation segment of AURKA is reported not to be in the fully active conformation in the absence of TPX2 (PDB: 1OL7).<sup>19</sup> The structure of phosphorylated TPX2-bound AURKA (PDB: 1OL5)<sup>19</sup> is highly similar to that of phosphorylated TPX2-unbound AURKA. The only difference is in the divergence of conformation of the activation loop between H280 – L293. (Figure 1.11) The crystal structures showed that the phosphorylated activation segment of AURKA in TPX2-bound state was pulled down compared to that in TPX2-unbound state. However, TPX2 is not always associated with AURKA. It is present for AURKA activity involved in mitotic spindle but not associated with control of the function of centrosome. This information is essentially related to the biochemical findings present in the next section.



**Figure 1.11.** Comparison of the conformations of the phosphorylated activation segment of AURKA (dotted purple) in two states: TPX2-bound (PDB: 1OL5) and TPX2-unbound (PDB: 1OL7).

Activity of AURKA is also further regulated by phosphatase PP1 and p53 tumour suppressor protein. AURKA has two PP1-binding sites. The AURKA mutants that are unable to bind PP1 were hyper-phosphorylated. Studies suggested that PP1 inhibits AURKA activity by dephosphorylating some crucial residues.<sup>48</sup> The p53 protein also appears to bind to the N-terminal non-catalytic domain of AURKA, suppressing the kinase activity.<sup>55</sup>

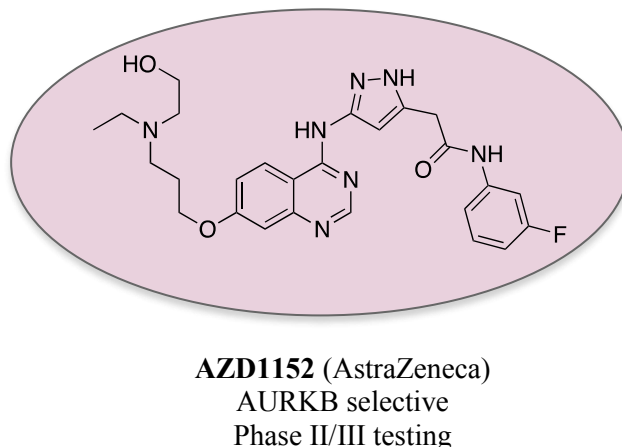
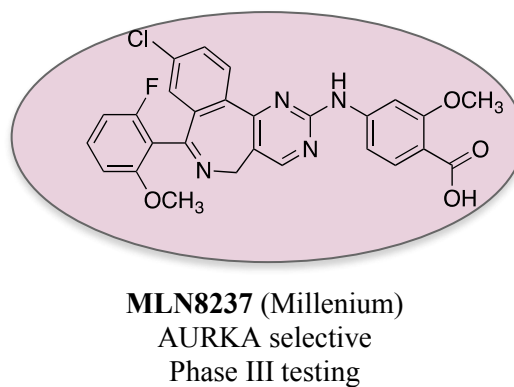
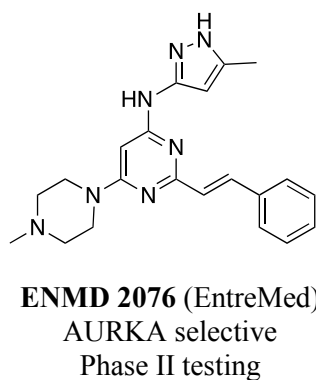
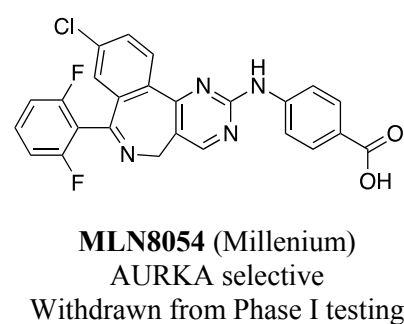
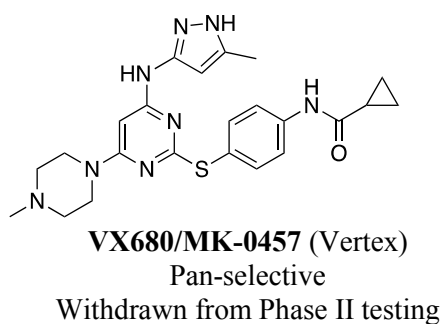
#### *AURKA Inhibitors*<sup>27, 56</sup>

Although there have been a number of Aurora kinase inhibitors entering phase I/II/III clinical trials, none have been approved as an anticancer drug. Most of them failed to pass through Phase I/II due to toxicity. In the review in 2016,<sup>27</sup> ten AURKA inhibitors were reported to yield some encouraging results in clinical trials. Among them, AURKA selective inhibitor MLN8237 (alisertib) is the most actively pursued agent, undergoing phase III testing in clinical trials.<sup>57</sup>

Some potent AURKA and AURKB inhibitors that made it to clinical trials are VX680,<sup>58</sup> AZD1152,<sup>59</sup> MLN8054,<sup>60</sup> MLN8237<sup>60</sup> and ENMD2076<sup>61</sup>. (Figure 1.12) Their IC<sub>50</sub> for Aurora kinases are summarised in Table 1.1.

**Table 1.1 – IC<sub>50</sub> of some current lead compounds for Aurora kinase inhibitions**

	IC <sub>50</sub> values in nM		
	AURKA	AURKB	AURKC
<b>VX680</b> <sup>58</sup>	0.6	18	4.6
<b>AZD1152</b> <sup>59</sup>	1369	0.36	n/a
<b>MLN8054</b> <sup>60</sup>	4	172	n/a
<b>MLN8237</b> <sup>60</sup>	1.2	397	n/a
<b>ENMD2076</b> <sup>61</sup>	14	350	-

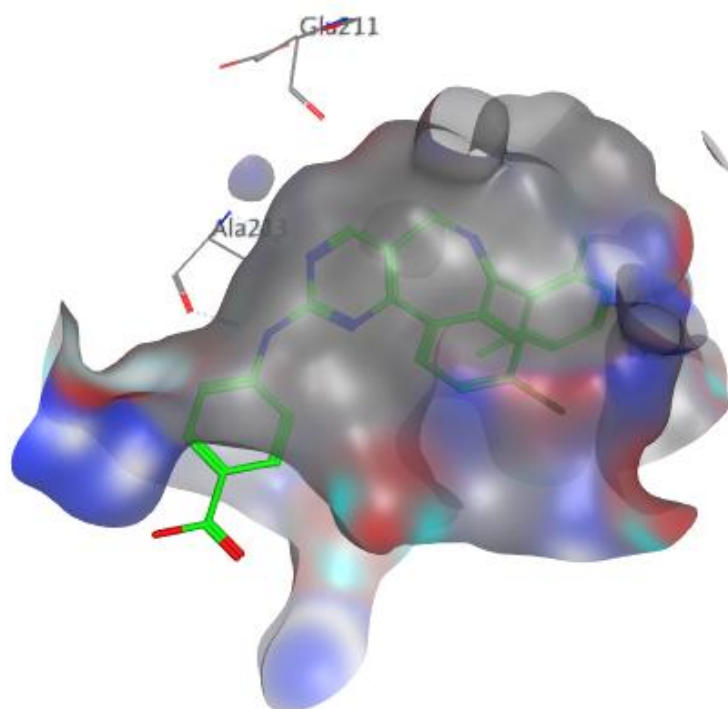


**Figure 1.12.** Chemical structures of some Aurora kinase inhibitors in clinical development.  
Compounds that have entered phase III (circled).

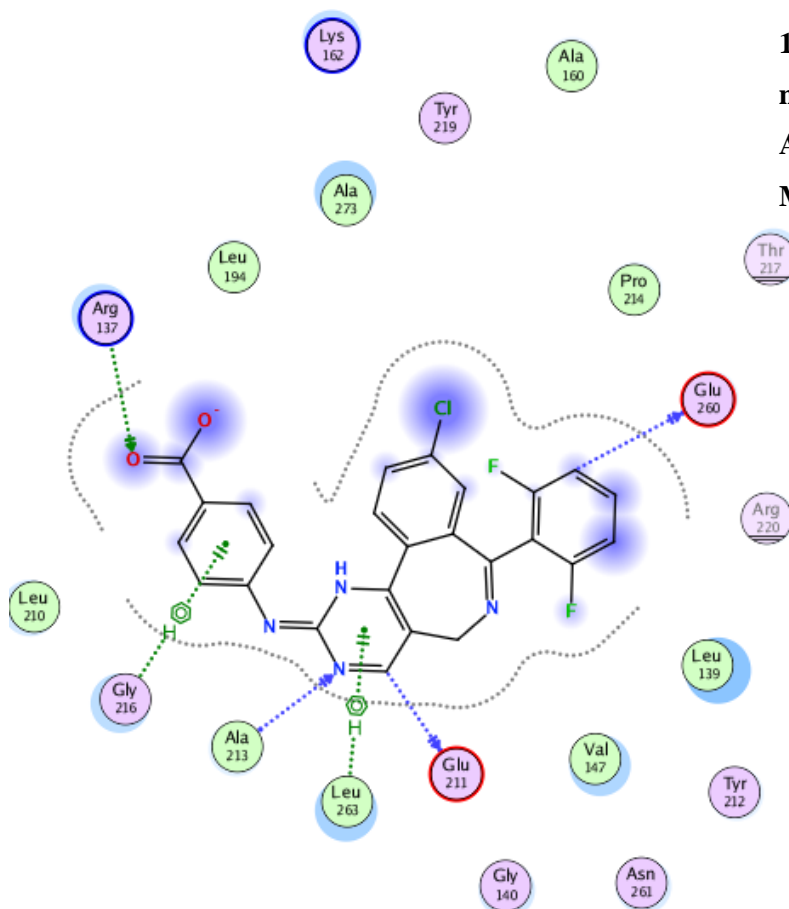
VX680 is a potent pan inhibitor of Aurora kinases. In *in vivo* leukaemia models VX680 was shown to reduce the tumour growth.<sup>62</sup> It also showed promising activity in patients with the T315I mutation who are resistant to BCR-ABL inhibitors such as Imatinib.<sup>63</sup> Unfortunately, VX680 was suspended in Phase I/II due to QTc prolongation (QT interval in the heart's electrical cycle represents electrical depolarization and repolarization of the left and right ventricles). MLN8054, developed by Millennium, is an ATP-competitive and reversible inhibitor, selective for AURKA. It exhibits a selectivity of >40-fold for AURKA over AURKB. Treatment of 1  $\mu$ M MLN8054 in human HCT-116 colorectal and PC-3 prostate tumour cells showed delay in G2/M progression.<sup>64</sup> However, MLN8054 terminated in Phase I due to CNS effects.<sup>65</sup> Its successor, MLN8237 seems to be more potent than MLN8054 and has less benzodiazepine-like effects. MLN8237 also showed high selectivity for AURKA (IC<sub>50</sub> of 1.2 nM) over AURKB (IC<sub>50</sub> of 397 nM), with over 200-fold selectivity in cell assay. Its anticancer activity was shown to be achieved by disrupting the assembly of the mitotic spindle apparatus and chromosomal segregation. It was also shown to inhibit tumour cell proliferation *in vitro* and tumour growth in solid tumour xenograft models.<sup>66</sup> A number of phase II clinical studies have been carried out on MLN8237 in patients with ovarian, fallopian tube and peritoneal carcinoma, acute myeloid leukaemia and high-grade myelodysplastic syndrome. Sixteen patients (52%) achieved stable disease with a mean duration of response of 2.86 months.<sup>67</sup> MLN8237 is the first oral AURKA selective inhibitor to enter Phase III trials and is currently being tested in patients with relapsed or refractory peripheral T-cell lymphoma (NCT01482962). ENMD2076, developed by Entremed, is another potent inhibitor that displays preferential inhibitory effects towards AURKA (IC<sub>50</sub> of 14 nM) to AURKB (IC<sub>50</sub> of 350 nM). This compound is also a potent inhibitor of Flt3 (IC<sub>50</sub> of 1.86 nM) and also inhibits 15 other oncogenic kinases with IC<sub>50</sub> values of less than 100nM, including VEGFR2/KDR, VEGFR3, FGFR1, FGFR2, and PDGFR $\alpha$  which are involved in angiogenesis.<sup>68</sup> It is currently undergoing phase II trials (NCT01639248). AZD1152 is a potent and selective inhibitor for AURKB, included here for possible comparison purposes. It is a pro-drug, which undergoes phosphatase-mediated cleavage to release barasertib-hQPA, the selective AURKB inhibitor.<sup>69</sup> It was shown to induce apoptosis of acute myeloid leukaemia and acute lymphoblastic leukemia cells.<sup>70</sup> In leukaemia clinical trials, AZD1152 has shown some adverse effects, including neutropenia and stomatitis/mucosal inflammation.<sup>71</sup>

At the molecular level, the binding positions of VX680 and MLN8054 inhibitors to the AURKA catalytic domain were shown to be around the hinge region, confirmed by crystal structures (PDB: 3E5A and PDB: 2X81<sup>72</sup>). There are currently no available co-crystal structures of MLN8237 and ENMD2076 bound to AURKA. However, due to the structural similarity of MLN8237 and MLN8054, it is predicted that MLN8237 may occupy similar

space. Figure 1.13 presents the interacting surface area and ligand interactions of MLN8054 and residues of the catalytic domain of human AURKA.



**1.13a. Surface interaction maps of MLN8054 in AURKA generated by MOE.** MLN8054 (C atoms in green). Hydrophobic area (grey), polar areas (red and cyan).



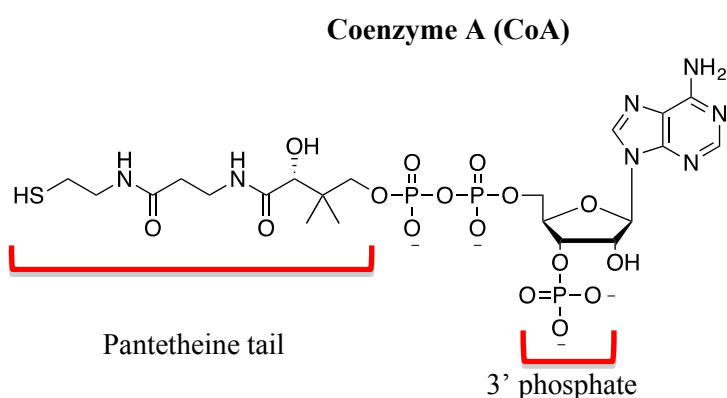
**1.13b. Ligand interaction maps of MLN8054 in AURKA generated by MOE.**

**Figure 1.13.** Interactions of MLN8054 with AURKA catalytic domain (PDB: 2X81).

Bayliss *et al.* in 2010 published the crystal structure of MLN8054 complexed to an AURKA mutant that mimicked AURKB (triple-point mutant: L215R, T217E and R220K to mimic the catalytic domain of AURKB). They reported that binding of MLN8054 to the AURKA mutant induced a change in the conformation of the activation loop, called DFG-up. The binding switched a polar pocket in the active site to hydrophobic, similar to what is observed when a molecule binds to AURKA and induces DFG-out conformation.<sup>29</sup>

In summary, oncogenic functions of AURKA are indicated in the process of proliferation, survival and invasion of tumour cells, thus it is a target for cancer therapy. Recently developed AURKA inhibitors have been shown to exhibit anticancer effects including suspension of tumour growth and induced apoptosis of cancer cells in clinical trials. However, the prime obstacle for the successful application of AURKA inhibitors remains to be its adverse effect of high toxicity. The reported number of toxicities were associated with AURKB inhibitors, including VX680 and AZD1152. The promising clinical results of AURKA selective inhibitor, MLN8237 encourage further efforts on the development of AURKA selective inhibitors. At the molecular level, most of the AURKA inhibitors have been shown to exploit the hinge region, the hydrophobic pockets and may reach the solvent accessible region. However, few explored the region below the ribose-binding pocket and phosphate binding region and activation loop. It is open for investigation that whether there will be any benefit in terms of inhibitory effect and selectivity from interactions with those regions.

#### 1.4. Biochemical findings<sup>73</sup>

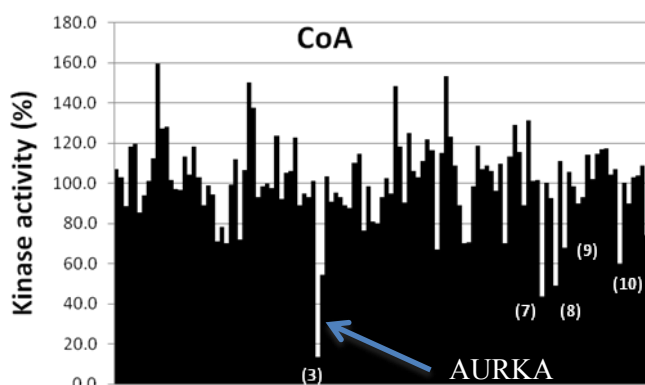


The following biochemical findings from Prof. Gout's group are the premise to this research project. The key findings are summarised below:

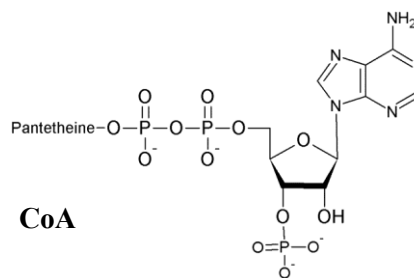
- (1) Preliminary screening of CoA's activity against a panel of 117 kinases showed that CoA exhibited striking selective activity.<sup>74</sup> (Figure 1.14) (Table 1.2) ADP, binding to

the hinge region of the catalytic domain, provides unselective inhibition of kinases. The addition of a pantheine tail to ADP moiety provides specificity in inhibitory effects towards 10/107 kinases. A further addition of the 3'phosphate group (as in CoA structure) or a combination of ADP and 3'phosphate increase the specificity towards AURKA, Src and YES1 kinases. These results show that both 3'phosphate and pantetheine are likely to be contributing to the specificity of CoA towards AURKA.

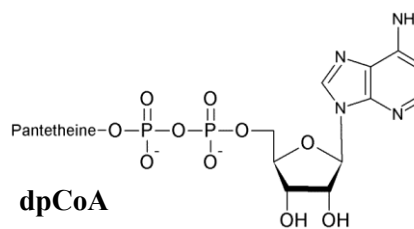
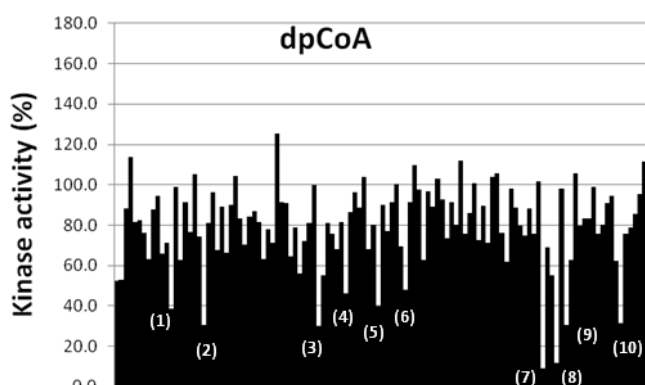




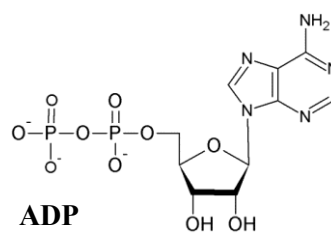
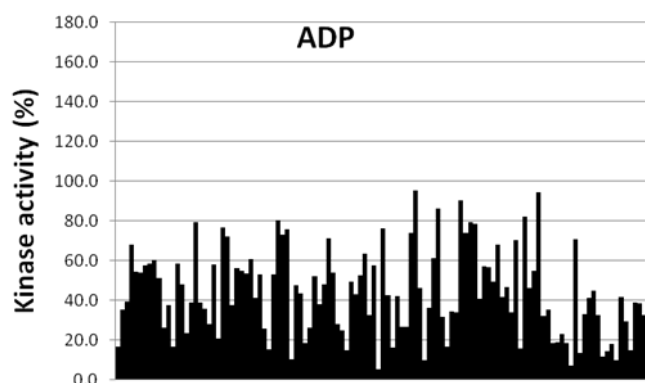
No of kinases inhibited >50 %



3/117



10/117



72/117

(1) ERK8; (2) S6K1; (3) AURKA; (4) MARK3; (5) CK2; (6) IKKb; (7) Src; (8) YES1; (9) BTK; (10) FGFR1

**Figure 1.14.** Selectivity towards kinase inhibition of CoA, dephospho-CoA (dpCoA) and ADP. [ADP] = 100  $\mu$ M, [CoA] = 100  $\mu$ M. The screening was done by International Centre for Kinase Profiling. The screening data for CoA were the average of 3 independent experiments. ADP and dpCOA were done once. <sup>74</sup>

**Table 1.2 – Screening assays of ATP, ADP, dpCoA and CoA against active and inactive AURKA**

Binding assay: Lanthascreen (Invitrogen), TR-FRET assay. Kinase activity was assayed by measuring incorporation of  $^{33}\text{P}$  ATP into myelin basic protein (Sigma). Purified protein (100ng) was incubated at room temperature for 30min in a total volume of 25 $\mu\text{l}$  containing 50mM HEPES pH 7.5, 150mM NaCl, 10mM MgCl<sub>2</sub>, 1mM EGTA, 0.05% Brij, 0.5mg/ml myelin basic protein, and 5 $\mu\text{M}$   $\gamma^{33}\text{P}$ -ATP (100-10,000dpm/pmol). The reaction was stopped by spotting the reaction mixture onto squares of P81 Whatman phosphocellulose ion exchange paper which were then immersed in 1.5% phosphoric acid. After washing twice in 1.5% phosphoric acid followed by two washes in distilled water, the papers were air dried and radioactivity was counted by a scintillation counter and recorded by Quantasart ver2.03.

Compound	AURKA (active) IC <sub>50</sub> ( $\mu\text{M}$ ) (Binding assay) <sup>†</sup>	AURKA (inactive) IC <sub>50</sub> ( $\mu\text{M}$ ) (Binding assay) <sup>†</sup>	AURKA IC <sub>50</sub> ( $\mu\text{M}$ ) (Activity assay) <sup>†</sup>
ATP	208	2990	-
ADP	117	1280	54.4 $\pm$ 9.3
dpCoA	~0.294	~7 x 10 <sup>17</sup>	16.9 $\pm$ 8.1 <sup>‡</sup>
CoA	~0.082	283	4.4 $\pm$ 2.2

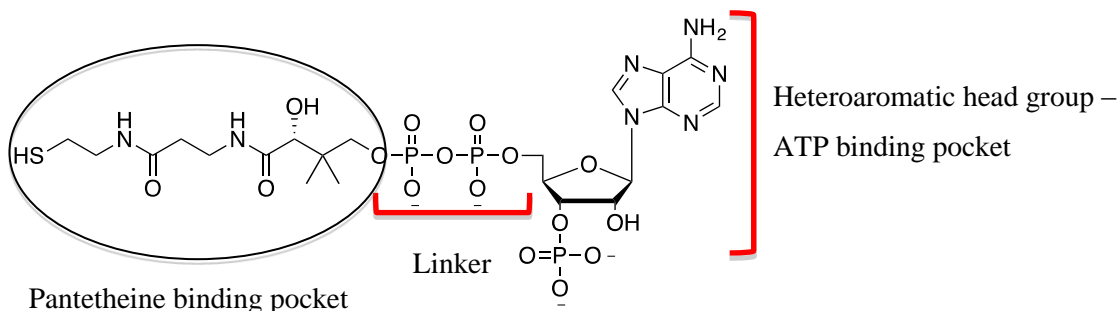
dpCoA: dephospho-CoA (dpCoA, removal of 3' PO<sub>4</sub><sup>2-</sup> group). Experiments done under [ATP] = 5  $\mu\text{M}$ . <sup>†</sup>Number of experiments (n) = 4; for dpCoA <sup>‡</sup>n = 3. Binding assays were done once.

(2) Dr. Y. Tsuchiya from Professor Gout's group showed that Coenzyme A (CoA) displayed a good inhibitory effect towards AURKA (IC<sub>50</sub> of 4.4  $\mu\text{M}$ ). (Table 1.2) Further ATP competitive assay experiments showed that CoA acted as an ATP-competitive inhibitor in the ATP binding site of AURKA. Table 1.2 showed that CoA also bound to active AURKA much better than to inactive AURKA. It also showed that dephospho-CoA (dpCoA), lacking the 3' phosphate, was also less potent than CoA (IC<sub>50</sub> of 17  $\mu\text{M}$ ) and did not bind to AURKA as strongly as CoA did. This suggests some interactions around the 3' phosphate binding region may be beneficial for AURKA inhibition.

(3) When TPX2 was included in assay buffer, no inhibition of CoA against AURKA was recorded. It suggests that TPX2 modified the catalytic domain to prevent CoA binding. This is confirmed when comparing the two AURKA structures. (Figure 1.11) Pre-incubating CoA and AURKA prior to the addition of TPX2 to the medium restored inhibitory activity of CoA. Furthermore, at 5  $\mu\text{M}$  of [ATP], in the presence of a reducing agent, DTT, IC<sub>50</sub> (CoA) is 47  $\mu\text{M}$  whereas in the absence of DTT, it is 4.4  $\mu\text{M}$ . This suggested that there are likely *two* steps for CoA binding to the catalytic domain of AURKA: (i) CoA binds to the hinge region of AURKA. The binding is weak intermolecular interactions/H-bonding; (ii) CoA rearranges itself in the active site and such arrangement makes it possible for a covalent bond to be formed.

(4) From the screening panel, CoA inhibited AURKA but not AURKB whereas dpCoA inhibited both. AURKA and AURKB share similar active site pocket with a difference in T217. This suggested some interactions between 3' phosphate of CoA and T217 region and such interactions contributed to the selectivity in inhibition. T217 has previously been shown to be important for AURKA sensitivity to the benzazepine-based Aurora inhibitor MLN8054 and the close analogue MLN8237. This residue was regarded as poorly accessible to a small molecule inhibitor.<sup>72</sup> Such hypotheses can be examined further through *in silico* experiments. (Chapter 2)

## 1.5. Thesis proposal



**Figure 1.15.** CoA scaffold for development of an AURKA selective inhibitor.

Based on the biochemical findings, coupled with *in silico* experiments and synthetic chemistry the project is focused on two areas:

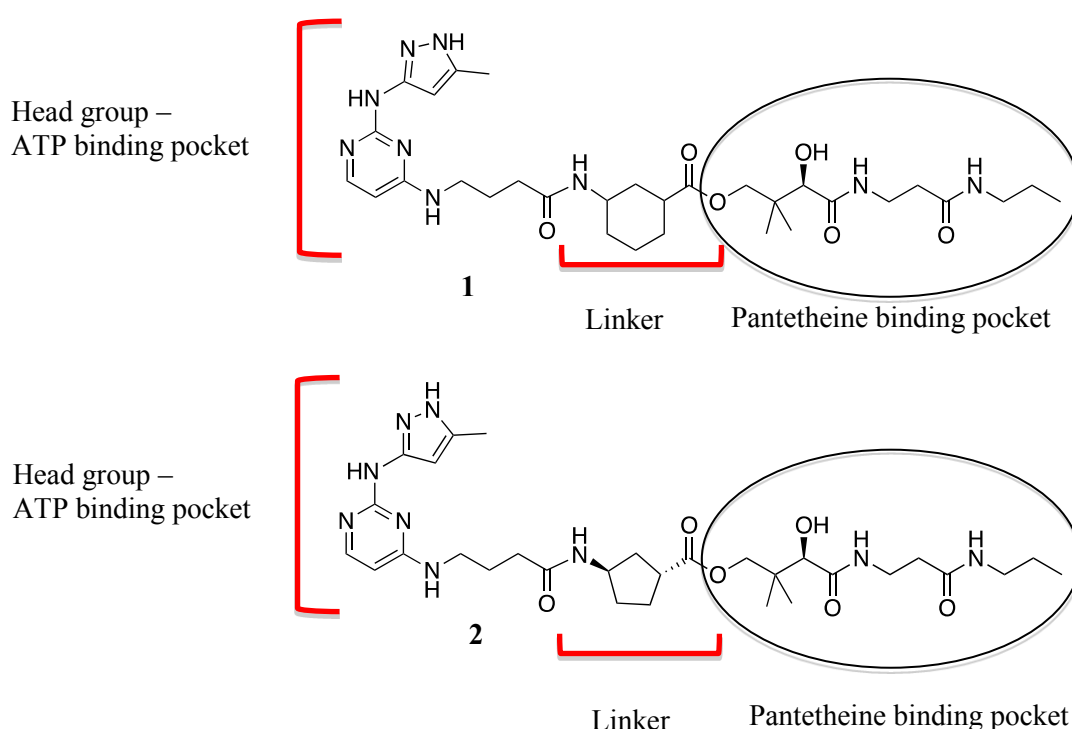
- (1) The possible binding mode of CoA in the ATP binding pocket of AURKA and an explanation of the selectivity of CoA towards AURKA.
- (2) CoA scaffold is utilised for the development of an AURKA selective inhibitor.

CoA molecular weight is well over 500 (molecular weight of MLN8237 is 518.92) and does not satisfy Lipinski's rule of 5.<sup>75</sup> It contains a phosphate and pyrophosphate linkage with the negative charge, making it unable to cross the cell membrane. Modification is needed to improve the drug-likeness and pharmacokinetic properties of CoA. As a result, the concept of "CoA scaffold" was established to design a selective inhibitor of AURKA. The CoA scaffold mentioned was defined by three components: a heteroaromatic head group, mimicking the adenosine group of CoA, a linker, mimicking the pyrophosphate of CoA and a pantetheine-mimicking tail. The initial ideas were to replace ADP moiety with a known hinge region-binding core and the pyrophosphate linker with a polar uncharged linker. The design would be assisted by computational modelling. The first design of the heteroaromatic head group was decided to follow that of VX680, with the aim to increase the binding at the hinge region compared to adenosine of CoA. It is described in detail in the first generation of compounds in Section 1.5.1. The synthesis of VX680 head group encountered problems and the initial computational predictions were found to contain some errors. This led to the decision to focus on verifying the biochemical findings and understanding the interactions of CoA in the AURKA catalytic domain more in detail before re-design of more inhibitory candidates. This "proof of concept" approach (Approach **B**) is the main content of this thesis.

### 1.5.1. Approach A – The first generation of compounds

As mentioned, in the first half of the project, much effort was spent on Approach A to design an AURKA selective inhibitor based on the idea of “CoA scaffold” and VX680. In approach A, the project was initially divided thus:

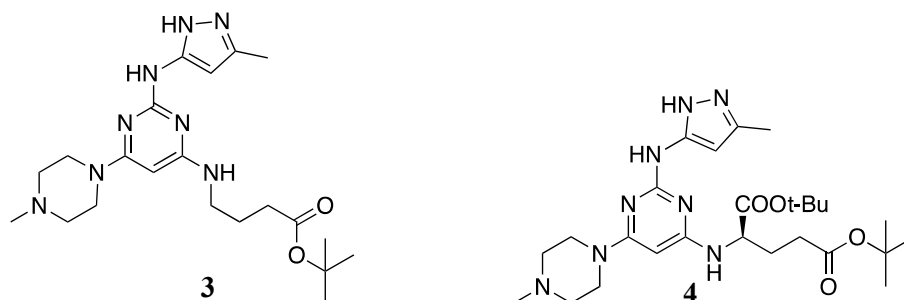
- (i) Computational analysis was done by our former collaborator, Dr. Lalitha Guruprasad. Unfortunately, this result was later found to be invalid. The synthetic target **2** was decided based on this work. (Section 2.4)
- (ii) Two synthetic structures (1<sup>st</sup> generation of compounds) were split into parts, which my colleague, Fiona Bellany and I, were responsible for preparing. (Figure 1.16) My colleague synthesised the head group whilst I focused on the synthesis of the linker and pantothenamide end. (Chapter 3)



**Figure 1.16.** Structures of the first generation of compounds, **1** and **2**.

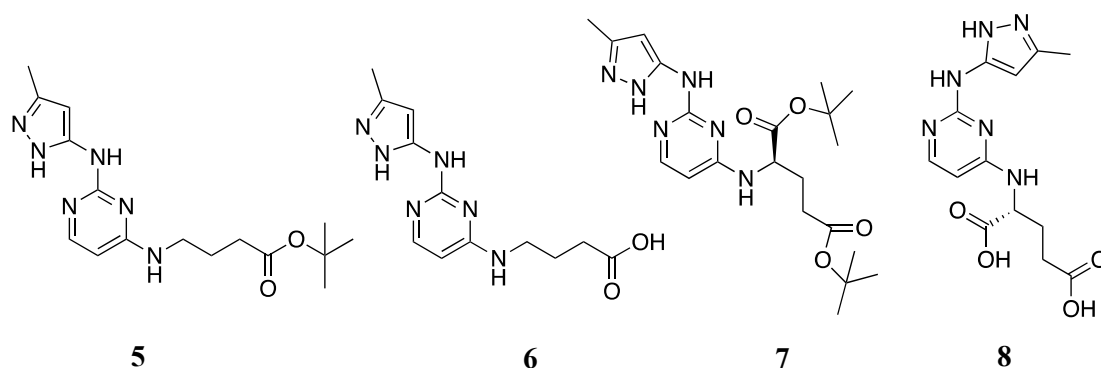
The main target of this approach was compound **2**. However, due to the high price of the linker, a test compound, compound **1** was set to be synthesised first to establish the synthetic route with the linker 3-aminocyclohexane-1-carboxylic acid available commercially as a mixture of *cis*- and *trans*-3-aminocyclohexane-1-carboxylic acid with >95% *cis*-3-aminocyclohexane-1-carboxylic acid. In this approach, as mentioned above, I focused on the synthesis of the linker and pantothenamide tail (i.e. pantetheine mimics) parts of the 1<sup>st</sup> generation of compounds (Figure 1.16) although some work was done on the investigation for synthesis of the head group (Figure 1.17), based on the preliminary docking of compound **1** and **2** and testing results of core structures prepared by my colleague, Fiona Bellany. (Table 1.3) Those synthetic results are

described in Chapter 3. The polar end –SH of the pantetheine tail of CoA was replaced by a –CH<sub>3</sub> group. It was suggested by the computational work done by Dr. Guruprasad that compound 1 and 2 would still fit well in the catalytic domain of AURKA and show significant potency. However, the biochemical results from Dr. Tsuchiya suggested that such change might reduce the potency of the compound. The initial plan was to establish a synthetic route to compounds 1 and 2, and then later a pantetheine end might be added to the compound to compare the effect of the –SH end.



**Figure 1.17.** Initial heteroaromatic analogues for AURKA inhibition.

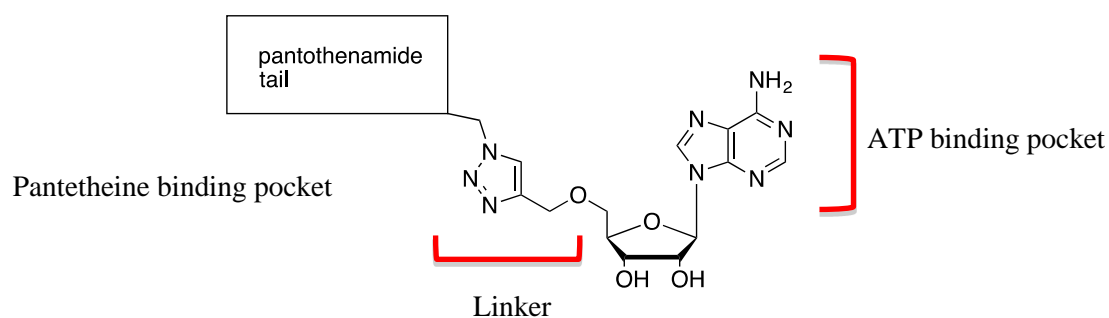
**Table 1.3 – Preliminary biological testing results of initial analogues for AURKA inhibition**



Compound	IC <sub>50</sub> values in μM	
	AURKA	AURKB
5	500	320
6	1137	1031
7	88	170
8	166	240

AURKA activity was assayed by a radiometric filter binding assay using myelin basic protein as a substrate. (See Table 1.2 legend for details). Experiments were carried out once (n = 1). Due to the poor inhibitory activity, exact IC<sub>50</sub> values were not pursued.

### 1.5.2. Approach B – The second generation of compounds



**Figure 1.18.** An example of the second generation of compounds

Having difficulty in designing and synthesising an AURKA selective inhibitor through Approach A, it was decided that more focus would be needed on synthesis of compounds to verify the biochemical and computational findings (“proof of concept”). As a consequence, the core of this thesis involves three aspects:

- (i) Computational study of the selectivity of CoA towards AURKA. (Chapter 2)
- (ii) Based on an understanding of the interactions between CoA and protein residues in the AURKA catalytic domain obtained, the *in silico* design of new inhibitor(s) was investigated. (Chapter 2)
- (iii) The organic synthesis of newly designed candidates. (Approach **B**) These designs serve as the “proof of concept” for the biochemical and computational findings and establish the foundation for further design of a selective inhibitor for AURKA. The derived synthetic plan employed an adenosine head, a triazole-ring as a linker together with the previously established synthesis of pantothenamide tail in Approach A. (Chapter 4)

## 2. Computational study and design

### 2.1. Backgrounds on proteins and methods of modelling

#### 2.1.1. Backgrounds

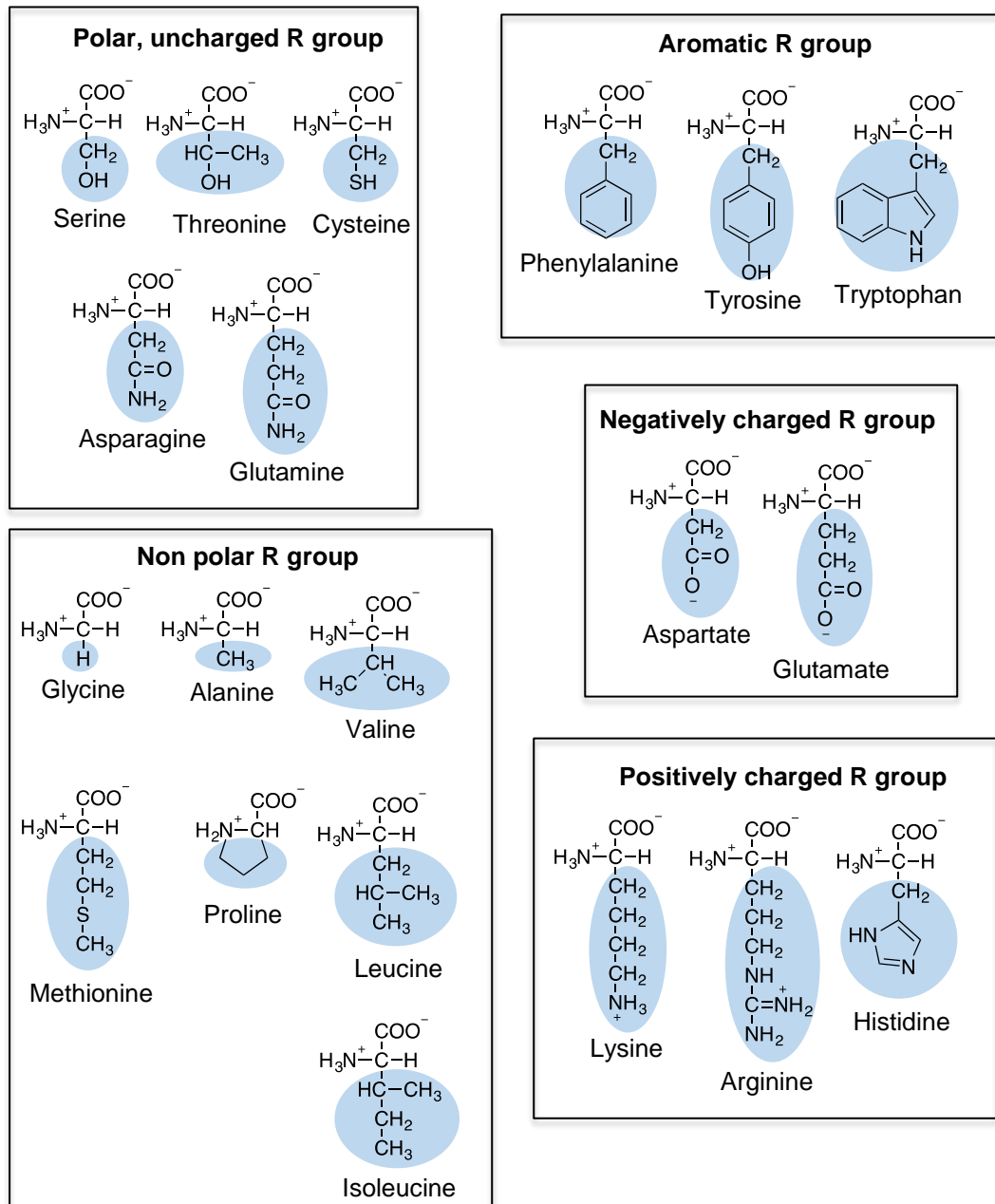
Understanding the structure-activity relationship of protein-ligand is essential in structure-based drug design. The striking selectivity of CoA towards AURKA made obtaining its structural binding information to AURKA desirable. It may also provide valuable structural information regarding the conformation of the activation loop of AURKA, which has not been well described.

The two most common experimental methods for structure determination are X-ray crystallography and NMR. Since effort on preparing crystal structures of CoA in AURKA has not been realised, the first step in modelling of CoA molecule in AURKA would be evaluation of an AURKA structure with a particular focus on the activation loop of AURKA. A background on protein structure and how to evaluate protein structure for modelling is given below. Once a structure is selected, docking and scoring of conformations of CoA in AURKA model will be carried out. The results will be analysed and compared with the previous biochemical results obtained by Dr. Tsuchiya, UCL to evaluate if they can explain those experimental observations. If the model is able to explain these biochemical results, it will be used to design compounds (CoA's analogues) to further assist the determination of CoA's conformation in AURKA and to explore possible areas for designing selective inhibitors of CoA. The model will be used to predict the behaviours of these designed compounds which will be verified later with biochemical experiments. If good agreements between the model and biochemical results are achieved, the model will continue to be used for design of an AURKA selective inhibitor. In circumstances where there is a lack of experimental information, theoretical methods such as homology modelling may be performed. It is only possible to build a confident homology model if there exist homologous proteins which have similar bioactive conformations of the desired structural region. The templates should share at least 40 % identical residues with the targets and the root-mean-square displacement (RMSD) values between final positions of  $C\alpha$  in the model and templates are less than 1.5 Å. If the sequences of two related proteins have 50% or more identical residues in an optimal alignment, the structures are likely to have similar conformations over more than 90% of the model.<sup>76</sup>

#### *Protein structure*<sup>77</sup>

Proteins are constructed from monomeric subunits, amino acids. The 20 common amino acids are  $\alpha$ -amino acids, having a carboxyl group and an amino group attached to the same C atom ( $C\alpha$ ). (Figure 2.1) Amino acids can be classified by R- group. The constituent amino acid residues in protein molecules are exclusively L stereoisomers.



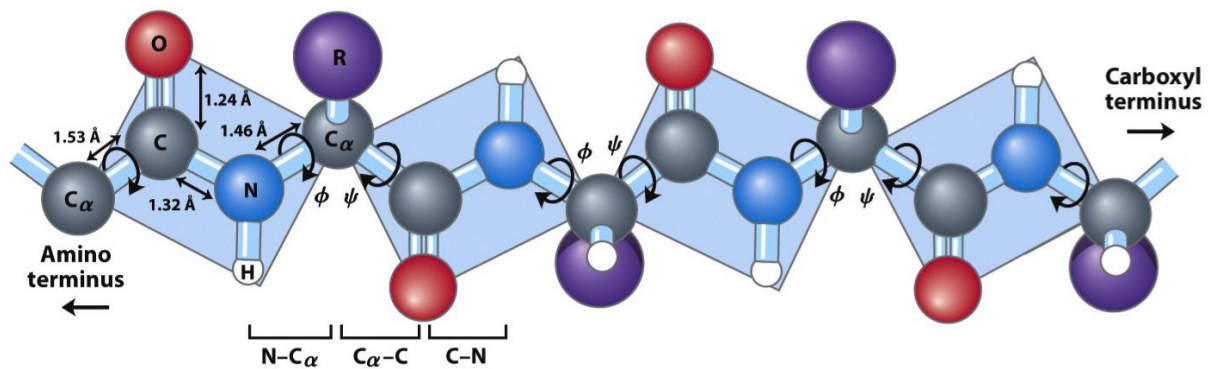


**Figure 2.1.** The 20 common amino acids of proteins. The structural formulae show the state of ionisation that would predominate at pH 7.0. The shaded portions are the sidechain R groups. Although the R group of histidine is shown uncharged, at pH 7.0 a small but significant fraction of the imidazole ring of the R group is positively charged.

Two amino acid residues can be linked together covalently through an amide bond to form a dipeptide. A polypeptide consists of many amino acids joined together. Some proteins such as lipoproteins, or glycoproteins may permanently contain chemical groups other than amino acid residues (e.g. lipids, sugars). A protein is commonly defined by four levels of structure. Its primary structure is a description of a sequence of amino acid residues linked together in a polypeptide chain. The spatial arrangement of amino acid residues that are adjacent in a

segment of a polypeptide constructs the secondary structure of a protein. As Pauling and Corey suggested in 1951,<sup>78</sup> a common protein secondary structure is a helical structure,  $\alpha$ -helix. In this arrangement, the polypeptide backbone is wound around an imaginary axis drawn longitudinally through the middle of the helix, making the R-groups protrude outward. The peptide plane is roughly parallel with the helix axis. The repeat unit is a single turn of the helix of 5.4 Å (3.6 residues). This arrangement maximises the use of internal hydrogen bonding. Another common protein secondary structure is  $\beta$  sheet, a more extended conformation of polypeptide chains. The adjacent polypeptide chains in a  $\beta$  sheet can have the same or opposite amino-to carboxyl orientations (parallel or anti-parallel respectively) and linked together by different hydrogen bonding patterns. The interstrand hydrogen bonds are essentially in-line in antiparallel  $\beta$  sheets and not in-line for the parallel ones. The repeat period is 6.5 Å for parallel conformation and 7 Å for antiparallel variant. The third type of secondary structure is  $\beta$  turns, which are the connecting elements that link successive runs of  $\alpha$  helices and  $\beta$  conformations. The tertiary structure of a protein refers to the overall 3-D arrangement of all atoms in a protein, which includes longer range aspects of amino acid sequence compared to the secondary structure. Globular proteins have a very compact structure. Most of the charged sidechains are on the surface or interacting with other polar groups. The sidechains of the protein core are usually non-polar or hydrophobic as found in insulin. Hydrophobic interactions are commonly found in the core residues of the proteins. Solvent molecules may be trapped inside cavities and form hydrogen bonding with protein residues. The folding of protein molecules can be explained as the consequences of the Second Law of Thermodynamics, which suggests that spontaneous processes always give rise to an increase of entropy of randomness, which in this case refers to the randomness of the solvent. The quaternary structure refers to the spatial arrangement of two or more polypeptide subunits of a protein. For example, insulin exists as dimers and hexamers.

The amide (peptide) ( $-\text{HNC}=\text{O}$ ) bond of a protein backbone is planar due to  $\pi$ -electron delocalisation. The carboxyl bond of the amide bond is 1.24 Å long, which is very short compared to the length of the carbon-oxygen single bond of 1.43 Å. The carbon-nitrogen bond of the amide bond of 1.33 Å is also shorter than the carbon-nitrogen single bond of 1.47 Å. Peptide conformation is defined by three dihedral angles called  $\phi$  (phi),  $\psi$  (psi) and  $\omega$  (omega), reflecting rotations about each of the three repeating bonds in the peptide backbone. The  $\phi$  angle describes rotation around N-C $\alpha$  bond and  $\psi$  angle describes rotation around C-C $\alpha$  bond. (Figure 2.2) Due to the nature of the peptide bond,  $\omega$  angle is restrained to value of 0° (*cis*-peptide) or 180° (*trans*-peptide). *Cis*-peptides are generally rare for steric reasons and may occur if the next residue is a proline (restricted rotation). Hence values of  $\omega$  usually offer little validation checks. The  $\phi$  and  $\psi$  angles are much less restricted and can take up values between -180° and +180°. However, due to steric hindrance, only specific combinations of the dihedral angles are allowed. The  $\phi/\psi$  diagram or Ramachandran plot<sup>79</sup> describes these relationships.



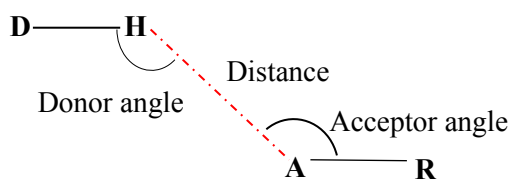
**Figure 2.2.** The planar peptide group. Three bond separate sequential C<sub>α</sub> atoms in a polypeptide chain. <sup>77</sup> (Reproduced from Figure 4-2a Reference 77. Used by Permission of the publisher.)

### *Protein-protein and protein-ligand interactions* <sup>80</sup>

The functions of many proteins are modulated by the reversible binding of other molecules called ligands and by protein-protein interactions. The two major factors determine these two interactions are geometrical and chemical complementarity between interacting parts. Geometry refers to the complementarity of the molecular surfaces of the two molecules/proteins involved and the chemical complementarity refers to the chemical interactions between the proteins and ligands/proteins. The generally reversible protein-ligand interactions and/or protein-protein interactions are governed by weak noncovalent interactions. These include the hydrophobic effect, hydrogen bonds, electrostatic interactions and van der Waals interactions.

In protein folding, the hydrophobic effect provides the thermodynamic drive for the overall structure of the proteins. The hydrophobic effect stabilises structures that exclude water from the interface because the entropy of the water is reduced when being exposed to hydrophobic surfaces. As in protein folding, the hydrophobic effect is essential for protein-protein interactions. Tsai *et al.* in 1997 reported from the data sets of 362 structurally non-redundant protein-protein interfaces and of 57 symmetry-related oligomeric interfaces that all protein-protein interfaces were hydrophobic. <sup>81</sup> When examining the hydrophobic effect in the data sets of 264 subunit-subunit, 59 receptor-ligand and 39 enzyme-inhibitor interfaces, they also found that the enzyme-inhibitor exhibited a stronger hydrophobic effect. This may be explained by the fact that the inhibitors which block the enzyme active sites are unlikely to dissociate as easily as the general protein-ligand bindings.

The hydrogen bond is another important class of molecular interaction in biology. It provides directionality and specificity to the protein folding, protein-protein and protein-ligand interactions. A hydrogen bond is defined as the electrostatic attraction between two polar groups, formed when a proton covalently bonded to a highly electronegative atom such as N or O (H-bond donor) experiences the electrostatic field of another electronegative atom such as N or O nearby (H-bond acceptor). <sup>82</sup> (Figure 2.3)



**Figure 2.3.** A schematic representation of the geometry of a hydrogen bond. D, donor atom; A, acceptor and H, hydrogen.

The amide and carbonyl groups of the peptide backbone in proteins as well as the polar functional groups (hydroxyls, amines, amides and carboxyls) of the sidechains can act as hydrogen bond donors and acceptors. In a study of various types of hydrogen bonds in protein-ligand complexes based on 4535 high quality protein-ligand complex structures by Liu *et al.* in 2008,<sup>83</sup> they showed that the optimal distance for a hydrogen bond between carbonyl oxygen and amide nitrogen was 2.5 – 3.5 Å, with a sharp peak at 2.9 Å. The angles at the donor and acceptor are greater than 90°. In particular, Baker and Hubbard in 1984 reported that the donor angle was between 130° and 170° and the acceptor angle was approximately 150°. (Figure 2.3) The bond is strongest when linear. In general, the strength of a single hydrogen bond ranges from 8 to 42 kJ.mol<sup>-1</sup>. In proteins, the strength of a single hydrogen bond is usually between 20 and 25 kJ.mol<sup>-1</sup>.<sup>84</sup> Apart from the hydrogen bond, other electrostatic interactions include the salt bridge. For example, the interaction between the –NH<sub>3</sub><sup>+</sup> group of the sidechain of lysine and the oxygen of a carboxylic acid group. Van der Waals interactions are another type of dipolar interactions. They usually refer to the weak and short-range attractive forces between uncharged molecules, as a result of the interaction of permanent or transient dipole moments. The distance over which van der Waals interactions occur is less than 7 Å between atom centres and is limited by electron repulsion range of 3.9 Å between atom centres according to the Pauli Exclusion Principle.

Covalent bonds are strong and lack the reversibility, and hence dynamics, of the weaker non-covalent bonds described above. Consequently, they are scarce in protein-protein and protein-ligand interactions. An example of this type of bond is the disulfide bond in immunoglobulins. Another example is the glutamate-lysine crosslinks which are important in crosslinking fibrin, the protein responsible for the stabilisation of blood clots. The covalent bond can be found in enzyme-inhibitor binding and is a characteristic of irreversible inhibitors. However, the majority of protein-protein and protein-ligand interactions are non-covalent.

*Evaluating protein structure for modelling*<sup>85, 86</sup>

Kleywegt summarised that a good model satisfies 6 requirements:

- Chemical requirements such as bond lengths and bond angles and steric constraints.

- Physical requirements such as favourable packing of the molecules in the crystal and sensible patterns of temperature factors.
- Crystallographic requirements: the model adequately explains the experimental data.
- Statistical requirement: the model is the best hypothesis to explain the data, with minimal assumptions.
- Protein structural requirement such as reasonable Ramachandran plot.
- Biological requirement: the model explains previous observations (e.g. with respect to activity, specificity, effect of mutations or inhibitors) and the model should enable you to make predictions that can be tested experimentally.

Crystal structures are the most common source of structural information for drug design. In order to obtain the best possible accuracy in modelling, the quality of protein crystal structure deposited in the database (e.g. RCSB PDB) needs to be assessed. An initial model built into an electron density map will need to go through model refinement and quality control to provide improvement in accuracy. When assessing a protein crystal structure, several refinement and quality control factors are taken into account. It is reviewed that a typically acceptable resolution is less than or equal to 2.5 Å.<sup>87</sup> A 2.5 Å *resolution* refers to the diffraction from sets of equivalent parallel planes of atoms spaced as closely as 2.5 Å in the unit cell. Other parameters to assess the precision of atomic positions in crystallography include R-factor and R-free. These are measures of the correlation between the model and the experimental data. The R-factor is calculated as follows:

$$R = \frac{\sum |F_{obs}| - |F_{calc}|}{\sum |F_{obs}|}$$

where  $|F_{obs}|$  is the measured intensity of a reflection in the diffraction pattern and  $|F_{calc}|$  is the intensity of the same reflection calculated from atomic model. Values of R-factors range from 0 to approximately 0.6. An R-factor value of 0.63 represents a totally random set of atoms and a perfect fit is represented by an R-factor of 0. The R-free value introduced by Brünger is a less biased way to examine the correctness of a model. R-free is calculated in similar way as R-factor but only uses a small randomly chosen set of intensities, test set (e.g. 5% of experimental data), which are not used during refinement. At any stage of refinement, R-free measures how well the atomic model predicts the test set that was not included in the refinement. R-free values are usually slightly higher than those of R-factor. Anderson suggested that in drug design, R-free value should be less than 0.28 and R-factor should be less than 0.25. Kleywegt, G. J. and Bergfors, T. *et al.* reported an example of the poor performance of the conventional R-factor as an indicator of the correctness of a model.<sup>88</sup> They described that the structure of cellular retinoic acid-binding protein type II (previously solved at 1.8 Å resolution) was intentionally traced backward and refined using data from 3 Å resolution. The conventional R-factor came

down as low as 0.214 which refers as ‘excellent’ but the R-free value came up at 0.67. They suggested that good data sets provided differences between R and R-free of approximately 0.02 and when resolution is low the difference may be as high as 0.05 - 0.08. Brünger also suggested that the precision of R-free values also depends on the size of the test set and ranges from about 0.5% at high resolution to 3 - 5% at low resolution.<sup>89</sup> However, R-free factor relates to the mean absolute phase error, hence local errors in a model may well go undetected if only R-free is used to assess the correctness of a model. Jones, T.A and Kleywegt, G. J. suggested that a low resolution model could not yield a precise model but that did not mean that one cannot produce an accurate model.<sup>85</sup> Precision is related to the level of detail and accuracy refers to how close to the truth something is. The knowledge of stereochemistry such as bond lengths, bond angles and violations of dihedral restraints may help increase the accuracy of a model in quality control. In particular the Ramachandran plot mentioned above would provide a description of the allowed residue conformations and outliers in a model. It usefully provides a quality assessment of a particular region of the protein, for example the activation loop of AURKA in this case. In other words, Ramachandran plot may be used to assess the stereochemical quality of a protein structure locally. Models that show poor clustering and many outliers are bound to be poor. Temperature factor, or B-factor is another parameter of quality control. Resolution presents the average uncertainty for all atoms whereas B-factor quantitates the uncertainty for each atom. It is a measure of the extent to which the atom is disordered within the structure, due to static or dynamic disorder within the crystal.<sup>86</sup> The B-factor of an atom  $i$  is calculated as follows:

$$B_i = 8\pi^2 U_i^2 \approx 79U_i^2$$

where  $U_i^2$  is the mean square displacement of the atom  $i$  from its resting position in  $\text{\AA}^2$ . For a value of  $B_i$  of  $79 \text{\AA}^2$ ,  $U_i^2$  is approximately  $1.0 \text{\AA}^2$  hence the RMSD of the atom  $i$  due to vibrations is  $1.0 \text{\AA}$ . Deviations from ideal bond lengths should be equal or less than  $0.015 \text{\AA}$  or  $3^\circ$  for bond angles. For macromolecule model, the best main chain deviations should be equal or less than  $0.4 \text{\AA}$  and sidechain deviations is less than  $1 \text{\AA}$  which translates into B-factor values of equal or less than  $15 \text{\AA}^2$  and  $79 \text{\AA}^2$  respectively. If there is no satisfactory protein structure available for modelling, a homology model from existing homologous proteins can be considered. Nevertheless a good model should satisfy the six requirements stated above, which include being able to explain biochemical experimental results and to make predictions that can be tested experimentally.

In this project, the main purpose of modelling is to probe the binding mode of CoA to AURKA, which hopefully provides an explanation of the selectivity of CoA towards AURKA kinase and to assist the design of new selective inhibitors for AURKA which may be based on the structure of CoA. I used bioinformatics techniques such as sequence and structural alignment and *in silico* mutant experiments to investigate the selectivity of CoA towards AURKA, and used a

docking programme in order to probe the binding mode of CoA to AURKA and to design analogues of CoA.

## **2.1.2. Materials and methods of modelling**

### ***2.1.2.1. Structural alignment***

Sequence alignment of protein sequences is a procedure of comparison of two or more protein sequences by searching for a series of individual characters or character patterns that are in the same order in the sequences. It provides useful structural, functional and evolutionary information for protein sequences. Alignment of two sequences is performed using the fast algorithm programmes FASTA<sup>90</sup> and BLAST (Basic Local Alignment Search Tool)<sup>91</sup>. Global multiple sequence alignment can be performed using CLUSTALW<sup>92</sup>. CLUSTALW is capable of providing an adequate alignment of a large number of related sequences and is a reliable indication of the domain structure of the examined sequences.

One of the “rules of thumb” is that if the sequences have more than 25% identical residues, they are likely to have a similar general folding pattern and if two proteins have more than 45% identical residues in their optimal alignment, they will have very similar structures. The patterns recognised in multiple sequence alignment may hold important information such as:

- Conserved sets of residues are likely to be part of the active site and may suggest some information about the structural-function relationship.
- The conservation patterns facilitate identification of other homologous sequences and classification of subfamilies within a set of homologues.

The multiple sequence alignment technique allows greater accuracy in alignment of distantly related sequences. Most structure-prediction methods are more reliable if based on a multiple sequence alignment. From 50,000 protein structures known, it is acknowledged that members of a protein family maintain the same basic folding pattern over ranges of sequence similarity from near identity down to well below 20% conservation. Families of related proteins tend to retain common folding patterns. Consequently, when comparing two or more protein structures, determination of residue-residue correspondences via structural superposition of two or more proteins is a more accurate method than sequence alignment for detecting homology and aligning the sequences of distantly related proteins.<sup>76</sup>

## **Method**

In this project, biological screening of CoA activity against a panel of kinases demonstrated that CoA is particularly selective towards AURKA (Kinase activity remaining is 14% at [CoA] of 100  $\mu$ M, IC<sub>50</sub> of 4.4  $\mu$ M). (Table 1.2) In order to investigate CoA selectivity for AURKA, structural alignment then comparison of AURKA to other kinases in the screening panel was carried out, particularly in the ATP binding region and the activation loop. Using the kinome as

a structural reference, a selection of kinases with different % Kinase activity remaining at [CoA] of 100  $\mu$ M in response to CoA was chosen for structural alignment. (Table 2.1) The kinome suggested that the closest structure to Aurora kinases was YANK kinases (YANK1, 2, 3), which are Ser/Thr kinases of the AGC family.<sup>12</sup> However, there is only one structure of YANK1 deposited in RCSB PDB, which is 4FR4<sup>93</sup> (of *Homo sapiens* species, Resolution 2.29 Å, R-factor 0.204, R-free 0.244). Moreover, there was no experimental data of % Kinase activity remaining at [CoA] of 100  $\mu$ M for YANK1. The next candidate which shared the most similar structure to Aurora kinases and was included in the screening panel was PKA of AGC family. Hence, PKA was selected for structural alignment and comparison. Other kinases in the screening panel were from different families such as TK and CMCG and CK1. They were selected based on the different values of % Kinase activity remaining at [CoA] of 100  $\mu$ M. Those kinases were Src and FGFR1 of TK family and CDK2 of CMCG family.

**Table 2.1 – Preliminary screening results of % kinase activity remaining at [CoA] and [ADP] of 100  $\mu$ M**

Kinase ( <i>homo sapiens</i> species)	% Kinase activity remaining at [ligand] of 100 $\mu$ M	
	CoA <sup>†</sup>	ADP <sup>‡</sup>
AURKA	14 $\pm$ 4	38
Src	44 $\pm$ 7	32
AURKB	55 $\pm$ 25	48
FGFR1	60 $\pm$ 10	42
CDK2	93 $\pm$ 9	27
PKA	99 $\pm$ 6	28

*In vitro* assays done by International Centre for Kinase Profiling. [ligand] = concentration of ligand; Ligand: CoA, ADP. <sup>†</sup> number of experiments (n) = 3; <sup>‡</sup> n = 1. [ATP] = 20  $\mu$ M except for AURKA: [ATP] = 5  $\mu$ M.

For 3-D structural comparison, crystal structures of the proteins are selected from the databases such as RCSB PDB, Kinase SARfari and/or Molecular Operating Environment programme (MOE) and aligned and visualised in MOE. The criteria for selection of the appropriate protein structure follows the six requirements for a good model mentioned above, including the evaluation of the refinement and quality-control factors such as the resolution, R-factor, R-free, B-factor and Ramachandran plot. In this project, the focus was on the comparison of the selected kinase structures around the highly conserved hinge region and the variable activation loop. In order to verify the quality of the protein structure of AURKA, virtual mutant experiments on T217 of the selected structure of AURKA was also carried out in MOE.



### 2.1.2.2. Docking programme <sup>94</sup>

In general, molecular docking programmes are used to model the interactions between a small molecule and a protein at the atomic level, which allow us to characterise the behaviour of small molecules in the binding site of target proteins as well as to elucidate fundamental biochemical processes. The theory of docking is to give a prediction of the protein-ligand interactions using computational methods. Fundamentally, the process of docking involves identifying a specific binding site, predicting the possible conformations of the ligand in the binding site of the protein by a search algorithm and then ranking these conformations via a scoring function. Ideally, the docking algorithms should be able to reproduce the experimental binding mode and the scoring function should also rank it highest among all generated conformations. The docking programme used in this project is GOLD 5.4.0, developed by Cambridge Crystallographic Data Centre (CCDC), to examine poses of non-covalently and covalently bound ligands. GOLD (Genetic Optimisation for Ligand Docking) is a genetic algorithm for docking flexible ligands into protein binding sites (predicting poses of ligands). It is generally used in structural prediction and in structural development of small molecule inhibitors. The GOLD scoring function, CHEMPLP (Piecewise Linear Potential), is an empirical scoring function optimised for pose prediction. In empirical scoring functions, the binding energy is a sum of several energy components, such as hydrogen bond, ionic interactions, hydrophobic effect and binding entropy. Each component is multiplied by a coefficient (weight) and then added together to give a final score (fitness score). Coefficients are obtained from regression analysis fitted to a test set of ligand-protein complexes with known binding affinities.

Piecewise Linear Potential (fPLP) is used to model the steric complementarity between protein and ligand, while for CHEMPLP additionally the distance- and angle-dependent hydrogen and metal bonding terms are considered (fchem-hb, fchem-cho, fchem-met). It is used for modelling the steric complementarity between a protein and a ligand as well as additionally considering the distance- and angle-dependent H-bonding terms.

$$fitness_{SPLP} = -(w_{PLP} \cdot f_{PLP} + W_{lig-clash} \cdot f_{lig-clash} + w_{lig-tors} \cdot f_{lig-tors} + f_{chem-cov} + w_{prot} \cdot f_{chem-prot} + w_{cons} \cdot f_{cons})$$

$$fitness_{CHEMPLP} = fitness_{SPLP} - (f_{chem-hb} + f_{chem-cho} + f_{chem-met})$$

$$fitness_{CHEMPLP} = fitness_{PLP} - (f_{chem-hb} + f_{chem-cho} + f_{chem-met})$$

*f<sub>lig-clash</sub>* : The internal score of the ligand consists of the heavy-atom clash potential

*f<sub>lig-tors</sub>* : The torsional potential

*wPLP : Weight of PLP contributions*

*Wlig-clash : Weight of ligand clash potential*

*wlig-tors : Weight of ligand torsion potential*

*wprot : Weight of ChemScore protein potential*

*wcons : Weight of constraint contributions*

The scoring function is capable of covalent docking (fchem-cov), considering flexible side-chains (fchem-prot) and explicit water molecules as well as handling constraints (fcons). The scoring calculations (fitness scores) are unit-less.

GOLD will dock each ligand several times, starting each time from a different random population of ligand orientations. The results of the different docking run are ranked by fitness score by CHEMPLP. By default the number of dockings to be performed on each ligand is 10. If the same pose is found in different docking runs, it is usually a strong indicator that it is the right pose. For a single ligand with default 10 dockings, GOLD will provide a list of ranked fitness scores. The largest fitness score of a ligand may indicate a higher chance of finding that pose in reality than other lower fitness scored poses of that ligand. However, for comparison between ligands, particularly ligands with significant different structures, the fitness scores are not a reliable measure. Validation test of GOLD on a set of 305 protein-ligand complexes in 2002 showed that it worked particularly well on kinases and isomerases with the success rates of 86% and 99% respectively on pdb files of 2.5 Å.<sup>95</sup> Astex scientists in 2007 have tested the reliability of GOLD docking in the prediction of protein-ligand interactions on their Diverse Set of 85 protein-ligand complexes of 11 kinases, 9 nuclear receptors, 5 serine proteases, and 3 members of the phosphodiesterase family.<sup>96</sup> They found that relatively unbiased protocols gave success rates of approximately more than 80% for redocking into native structures of large site (10 Å) and it is possible to get success rates of over 90% with some protocols. FLEXX is another automatic docking tool used for large data sets but the success rates reported have been lower than GOLD.<sup>97</sup> In a recent validation test on Astex Diverse Set of 85 protein-ligand complexes, CHEMPLP was found to give the highest average success rates (81%) for both pose prediction and virtual screening experiments against the other three GOLD scoring functions: GoldScore (69%), ChemScore (76%) and ASP (72%) and was set as the default scoring function in GOLD.<sup>98</sup>

## **Method**

Based on the literature statistics, in this project, in order to probe the conformation of CoA inside AURKA, GOLD CHEMPLP was used for docking of CoA into the appropriate AURKA protein structure. The first step was to test if the docking function, CHEMPLP, was able to reproduce the conformation of ADP in AURKA, which was available in the crystal structure. If

it was able to predict accurately the conformation of ADP in AURKA, it would be then used to dock CoA in AURKA. The ranked docking conformations by GOLD CHEMPLP were then visualised and analysed in MOE to select the best predicted conformation of CoA in AURKA. GOLD docking programme was also used to predict the conformations of analogues of CoA, which were designed to assist the evaluation of the conformation of CoA in AURKA and to develop a new AURKA selective inhibitor.

#### *Examination of appropriate kinase protein structure of AURKA*

All the available crystal structures of the examined kinase protein are downloaded through MOE database, and/or RCSB PDB and examined by superimposing them together (same spatial coordinations). In this project, 72 protein structures of Aurora kinases were downloaded from MOE database.

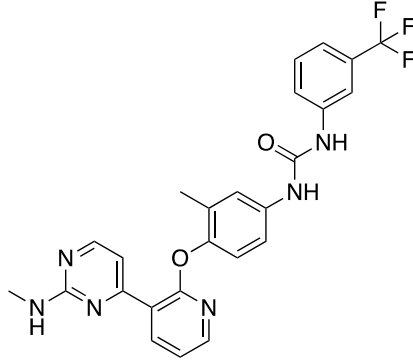
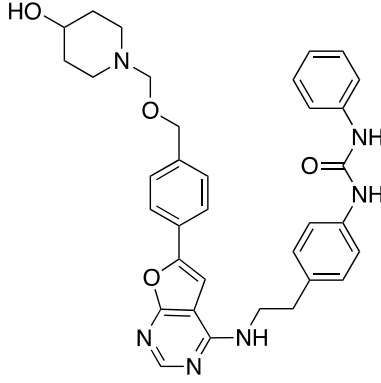
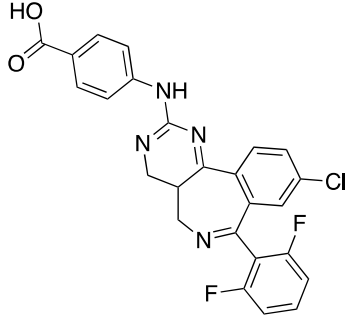
Based on the objectives of the project, a set of criteria was set to select the appropriate crystal structure. CHEMPLP was re-validated by docking the original ligand back into the protein crystal structure to see if the same result as X-ray crystal structure can be obtained. In this project, there were three criteria to select an appropriate AURKA kinase crystal structure:

- (i) It is in the phosphorylated form of AURKA where T287 and T288 were phosphorylated and TPX2 was not bound. DFG-in conformation of classic kinase binding site.
- (ii) The presence of C290 residue of the activation loop in the structure.
- (iii) Examination of the quality of the crystal structures using global parameters such as resolution, R-factor, R-free factor and local parameters such as Ramachandran plot and B-factor.

#### *Validation of docking programme and scoring function*

GOLD CHEMPLP programme was used to redock the co-crystallised molecules back into the original protein structures (in this project, AURKA); for example, molecule AK8 in 3EFW (organism: *Homo sapiens*),<sup>99</sup> YPH in 4JBP (*Homo sapiens*)<sup>100</sup> and ZZL in 2WTV (*Homo sapiens*)<sup>29</sup>. The molecules chosen were of the inhibitors with highest molecular weights in RCSB database, which are likely to expand a larger space compared to those with lower molecular weights. The reason for this was because CoA is of high molecular weight (compared to the general small molecule ligands with MW of less than 500 Da) and occupies a significant amount of space within the ATP binding site. The use of different expression systems was to test the general ability of ChemPLP to reproduce the X-ray conformations. (Table 2.2) ChemPLP should successfully produce the binding conformations of these molecules in the binding sites of the proteins judged by the largest scored conformations.

**Table 2.2 - Some current crystallised inhibitors of AURKA and AURKB.**

PDB	Ligand structure	IC <sub>50</sub>	IC <sub>50</sub>	GOLD highest score
		[AURKA] nM	[AURKB] nM	
3EFW		1	1	89.8
4JBP		41	17	113.9
2WTV		4	172	87.5

#### *Setting up protein and ligand for docking*

The ligands such as CoA were set up in their lowest energy conformation with hydrogen atoms added and were saved in .mol2 file format. For the protein files, water molecules were removed as they were not in the catalytic domain (binding site) and hydrogen atoms were added. The binding site was the catalytic domain of the kinase (AURKA), extending from the hinge region to the activation loop. Protein flexibility was not addressed in this test set.

Non-covalent and covalent dockings of the investigated ligands such as CoA or any designed compounds in Section 2.3 were done using GOLD CHEMPLP. Covalent docking is possible if you specify which ligand atom is bonded to which protein atom. The docking results were visualised and analysed in MOE.

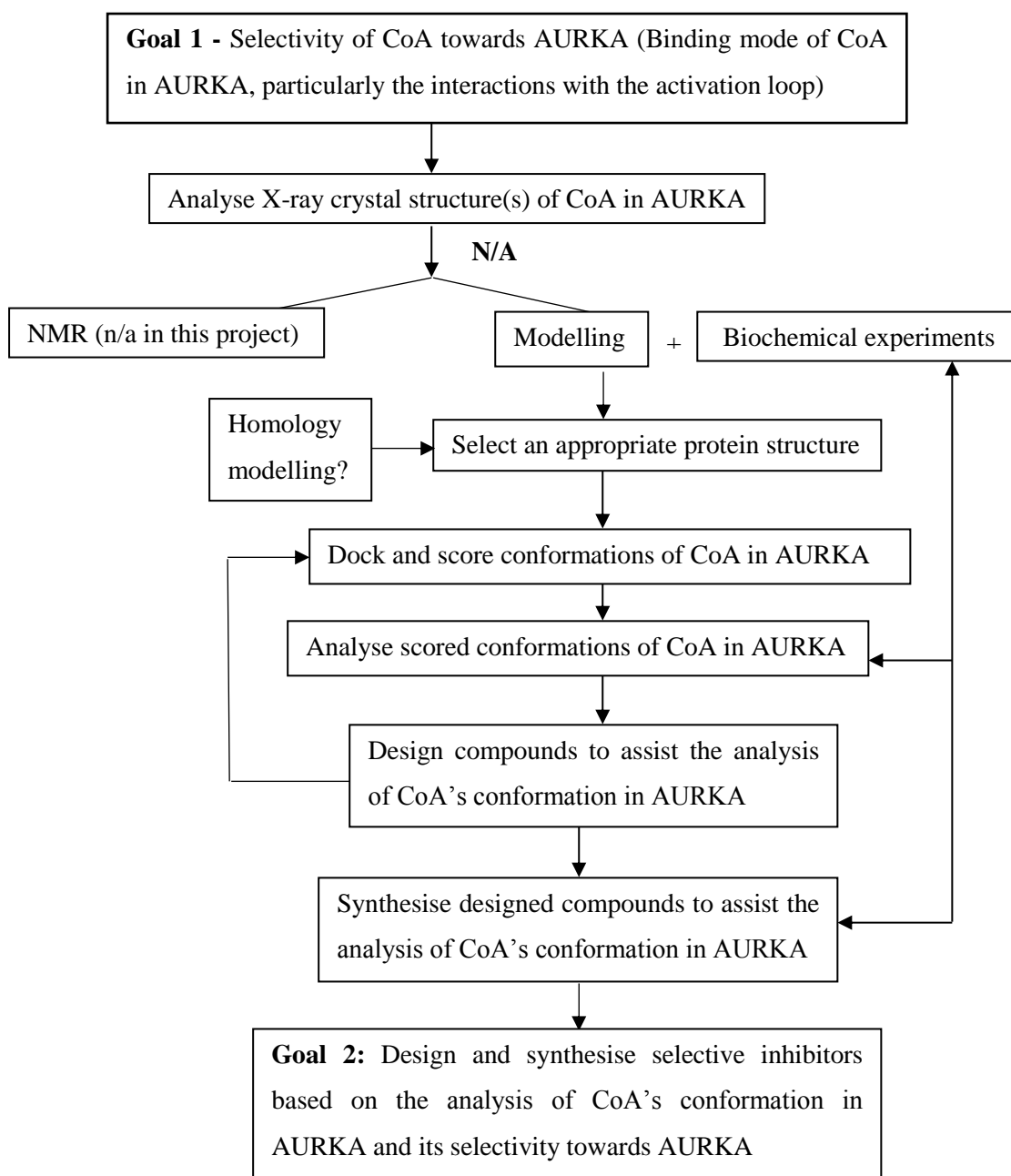
### *Analysing the docking results*

The ranked conformations of the ligand in the protein-ligand complexes by CHEMPLP were visualised and analysed in MOE. The filters to detect suspect entries included:

- Binding of adenosine head group outside of the hinge region
- Presence of severe clashes between protein and ligand (CoA)
- Unlikely geometry of the ligand (CoA)

### *In summary,*

In order to explain the selectivity of CoA towards AURKA in the screening panel (the structural-activity relationship), a set of kinases including, human AURKA, AURKB, Src, FGFR1, CDK2 and PKA, which exhibited different values of % Kinase activity remaining at [CoA] of 100  $\mu$ M, was selected. The methods used included sequence and structural alignment and *in-silico* mutant experiment using CLUSTALW and MOE respectively. In order to probe the binding mode of CoA to AURKA, an evaluation of the appropriate AURKA structure was carried out followed by docking of CoA and its analogues into that structure using GOLD CHEMPLP. The planned process of structure-based drug design in this project is summarised below.



**Figure 2.4.** The planned process of structure-based drug design for an AURKA selective inhibitor based on CoA. (n/a: not available)

## 2.2. Backgrounds on computational tools

### 2.2.1. PROCHECK for Ramachandran plot <sup>101</sup>

For a pdb file, a Ramachandran plot ( $\phi$  angle in x-axis against  $\psi$  angle in y-axis) is available calculated with PROCHECK programme. Morris *et al.* in 1992 <sup>86</sup> defined that the colouring/shading on the plot represents the different regions with the darkest areas usually refer to the "core" regions, which represent the most favourable combinations of  $\phi/\psi$  values. Ideally, one would hope to have over 90% of the residues in these "core" regions.

### 2.2.2. MOE for molecular visualisation and amino acid alignment and modification <sup>102</sup>

MOE (Molecular Operating Environment) programme, version 2015.10, developed by Chemical Computing Group (CCG), is a comprehensive molecular modelling package. It was used to visualise the docking solutions by GOLD and to perform structural alignment and modification (*in silico* mutant experiment) of different protein kinases. (Such experiments are presented in Section 2.2.2)

MOE platform allows analysis of further docking poses of ligands inside the specified docking site of the receptors, including contact statistics, electrostatic and interacting surface area map. Other functions of MOE, which were not used are not mentioned.

### 2.2.3. Other bioinformatics tools

#### LIGPLOT <sup>103</sup>

LIGPLOT, now superseded by new version, LigPlot+, is a graphical front-end to the LIGPLOT and DIMPLOT programs written by Andrew Wallace and Roman Laskowski.

This programme generates schematic diagrams of protein-ligand interactions for a given PDB file. Similar function is available in MOE Ligand Interactions command.

#### ChEMBL database <sup>104</sup>

ChEMBL is a service part of the European Molecular Biology Laboratory (EMBL), providing an open data resource for binding, functional and ADMET (Absorption Distribution Metabolism Excretion and Toxicological) bioactivity data.

#### PDBsum <sup>105</sup>

PDBsum is a pictorial database that provides an overview of the contents of each 3D structure deposited in the Protein Data Bank (PDB).

It shows the molecule(s) that make up the structure (ie protein chains, DNA, ligands and metal ions) and schematic diagrams of their interactions. Extensive use is made of the freely available RasMol molecular graphics program to view the molecules and their interactions in 3D.

#### SAS (Sequence Annotated by Structure) <sup>106</sup>

When searching for other pdb files of a particular protein with a known good pdb file, SAS is used. SAS is a tool for applying structural information to a given protein sequence. It uses FASTA to scan a given protein sequence against all the proteins of known 3D structure in the

Protein Data Bank (PDB). The resultant multiple alignment can be coloured according to different structural features and the matching 3D structures can be superimposed and viewed in MOE/PyMOL.

#### RCSB PDB <sup>107</sup>

RCSB PDB is an information portal for 117882 biological macromolecular structures. This archive contains information about the 3D structures of large biological molecules, including proteins and nucleic acids of many different species.

#### CHEMDRAW <sup>108</sup>

CHEMDRAW/CHEMOFFICE Professional is a complete drawing tool used to create scientifically intelligent drawings for molecules (2-D and 3-D) and to provide information on the chemical properties of the molecules such as molecular weight and clogP value.

*Octanol-water partition coefficient logP* is used as a measure of molecular hydrophobicity. It is defined as the ratio of concentrations of the unionised compound between organic (e.g. octanol) and aqueous phases at equilibrium. . Its representing equation is as follows:

$$\log P = \left( \frac{[solute]_{octanol}^{unionised}}{[solute]_{water}^{ionised}} \right)$$

Hydrophobicity affects drug absorption, bioavailability, hydrophobic drug-receptor interactions, metabolism of molecules, as well as their toxicity. In the simplest term, the logP values are linked to the solubility in water of the molecules. A high value of logP is related to low hydrophilicity of the molecule in question.

The *calculated logP (clogP)* value in CHEMDRAW is based on a pool of fragments predefined in the logP calculator. <sup>109</sup> The clogP value of a molecule is the sum of the fragment values that are present in the molecule.

### **2.3. Results and Discussion on the selectivity of CoA towards AURKA**

#### **2.3.1. Structural alignment of selected protein kinases for structural comparison**

##### ***2.3.1.1. Selection of protein kinase structures for alignment and docking***

###### *Selection of AURKA protein structures*

Firstly the crystal structures of AURKA and AURKB from all species in the protein database was downloaded through MOE Protein Kinase Database (Appendix 2.1). There are 72 structures. All the crystal structures are structurally aligned in MOE for the purpose of structural comparison.

The catalytic domain of AURKA is large and defined by five features: (Figure 2.5)

- (i) Hinge region (yellow): residues E211 – A213, extended to T217 region where many small inhibitors (e.g. MLN54, MLN8237 and VX680) bind to through H-bonding.
- (ii) Flexible chain (green): residues L139 – V147 and A160 – K162



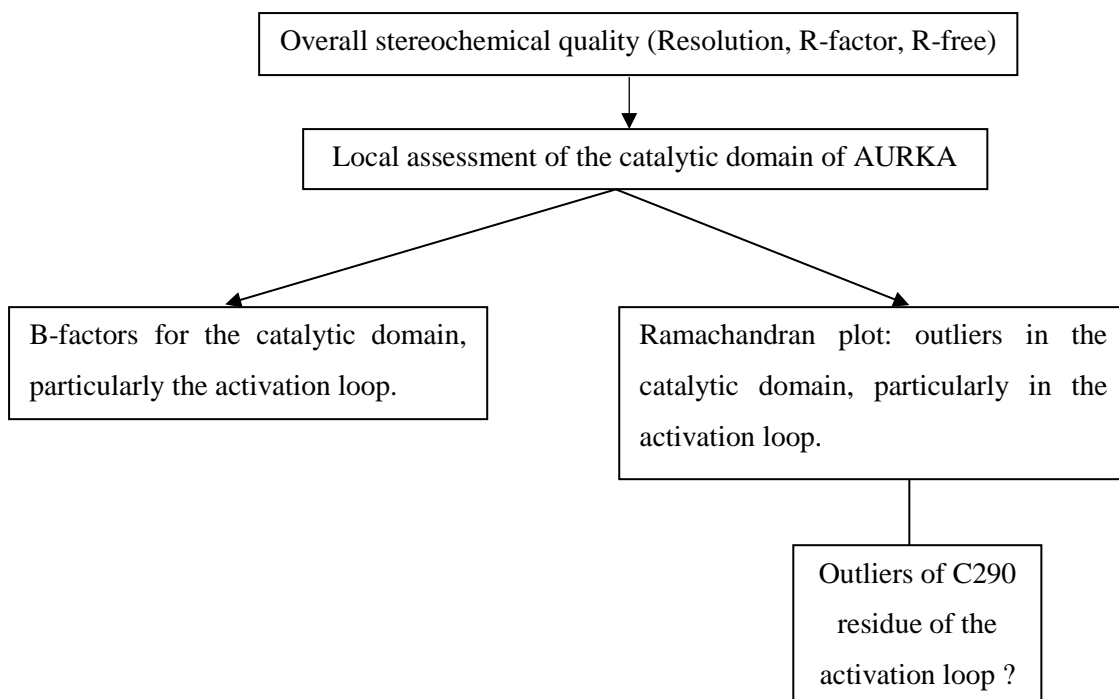
- (iii) Back loop (purple): residues L178 – Q185
- (iv) Front “loop”/surface: residues (red) H254 – L263 extended to residues A273 – G276 (cyan), forming a front surface.
- (v) Activation loop (residues in pink) between ‘red’ and ‘cyan’ region: W277 – D294.



**Figure 2.5.** Superimposed AURKA and AURKB crystal structures (72 structures from MOE database, Appendix 2.1). Hinge region (yellow), flexible chain (green), back loop (purple), front “loop”/surface (red and cyan), activation loop (pink, different conformations: circled).

For the original 72 structures of crystallised Aurora A and B kinase structures of all species (human and non-human) available, the first criteria of selection applied was that the crystal structure of AURKA was in the active form where T287 and T288 were phosphorylated and TPX2 was not bound. This criteria was derived from the biochemical experimental results report by Dr. Tsuchiya (Section 1.4, key findings (2) and (3)) where CoA was reported to bind to AURKA in its phosphorylated form without TPX2. This eliminated 63 structures. The 9 structures that met the criteria were 1OL7,<sup>19</sup> 4BN1,<sup>110</sup> 2WTV.A, 2WTV.B and 2WTV.D<sup>29</sup> for AURKA, and 4C2V.A, 4C2V.B,<sup>111</sup> 4C2W.A, 4C2W.B<sup>111</sup> for AURKB. The second criteria was the presence of C290 residue of the activation loop in the structure. As can be seen from Figure 2.6 this activation loop was missing or adopted variable conformations due to its flexibility in many pdb protein files making them impossible to be the candidates for docking. (Figure 2.5, circled) From the biochemical findings (Section 1.4, key finding (3)), it was found that at 5  $\mu\text{M}$  of [ATP], in the presence of a reducing agent, DTT,  $\text{IC}_{50}$  (CoA) is 47  $\mu\text{M}$  whereas in the absence of DTT, it is 4.4  $\mu\text{M}$ . It was suspected that the sidechain –SH of C290 residue may form some interactions with the end –SH or CoA. Therefore a structure containing C290

residue was desirable. Out of these 9 structures, three of them (2WTV.A, 2WTV.B and 2WTV.D) had C290 modified, forming disulfide bond with –SCH<sub>2</sub>CH<sub>2</sub>OH. As a result, these three structures were excluded, leaving two structures, 1OL7 and 4BN1 left. The comparison of the quality of 1OL7 and 4BN1 was carried out in the following order:



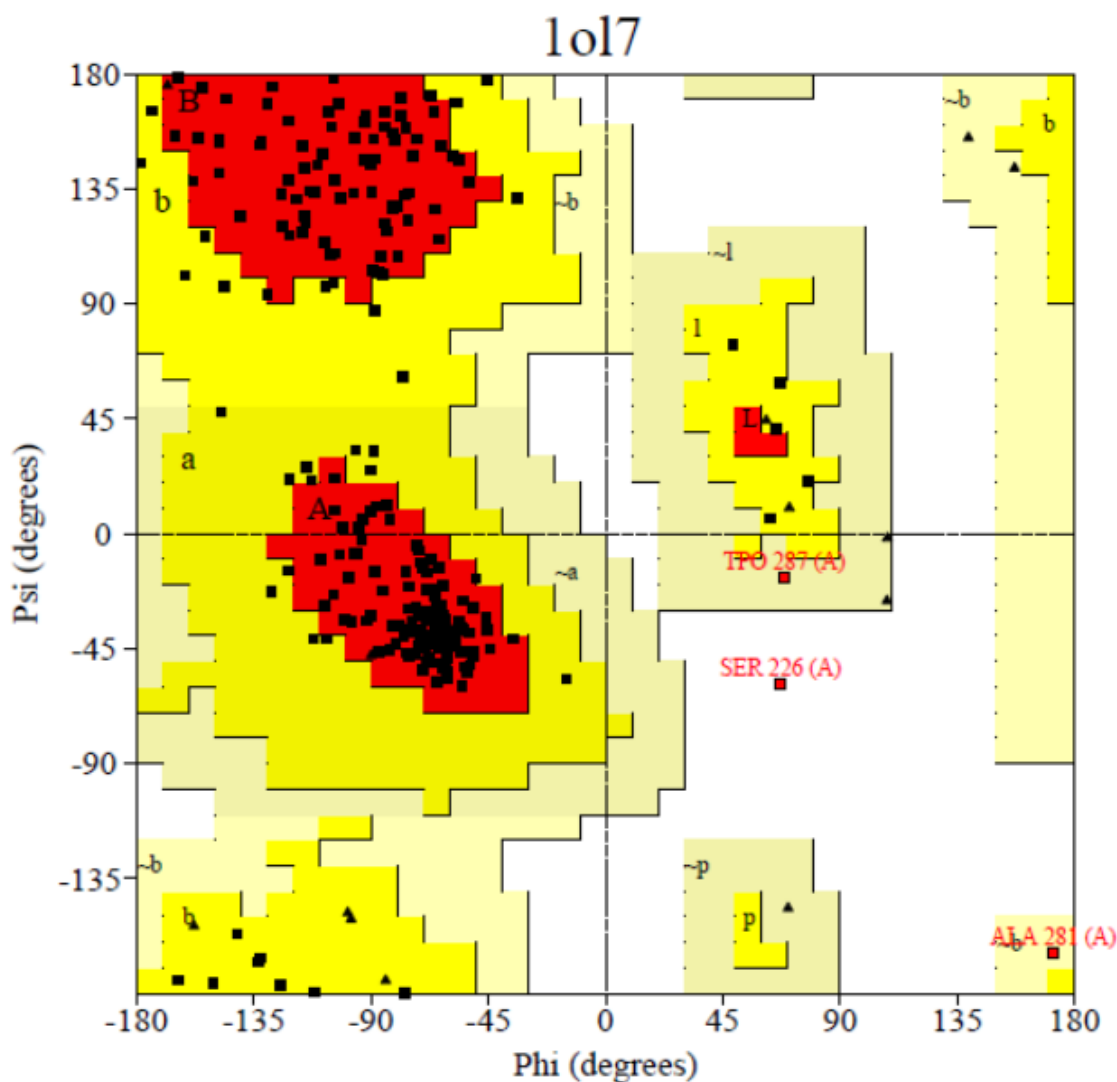
Checking the overall global parameters for structural validation, 4BN1 was slightly better than 1OL7 in term of quality. The resolution of the two pdb files, are 2.75 Å for 1OL7 and 2.50 Å for 4BN1). For R-factor value, that of 1OL7 is 0.257 and that of 4BN1 is 0.195. For R-free value, that of 1OL7 is 0.296 (free test size of 5%) and that of 4BN1 is 0.244 (free test size of 4.8%). (Table 2.3)

**Table 2.3 – Comparison of global parameters to assess quality of protein structure of AURKA between 1OL7 and 4BN1**

Parameter	1OL7	4BN1
Resolution/ Å	2.75	2.50
R-factor	0.257	0.195
R-free	0.296	0.244

Both free test sizes are similar and on the high end. It was suggested that the R-factor should be less than 0.25 and R-free value should be below 0.28 for drug design. Those values of 1OL7 were slightly out of range. However, a local investigation in the 3-D alignment of 1OL7 and 4BN1 in MOE showed that the conformations of amino acids in the five regions of the catalytic domain of AURKA were mostly the same. Variations were found in the location of the sidechains of residues in the activation loop, the area of interest. One particular important

exception was in C290 of the activation loop. In 1OL7 model, the sidechain –SH was reported to be missing whereas in 4BN1 model, it was built in a way that the –SH was pointing outwards of the activation loop. Ramachandran plots of 1OL7 and 4BN1 were examined for a local assessment of the stereochemical quality of the activation loop. (Figure 2.6 and Figure 2.7)

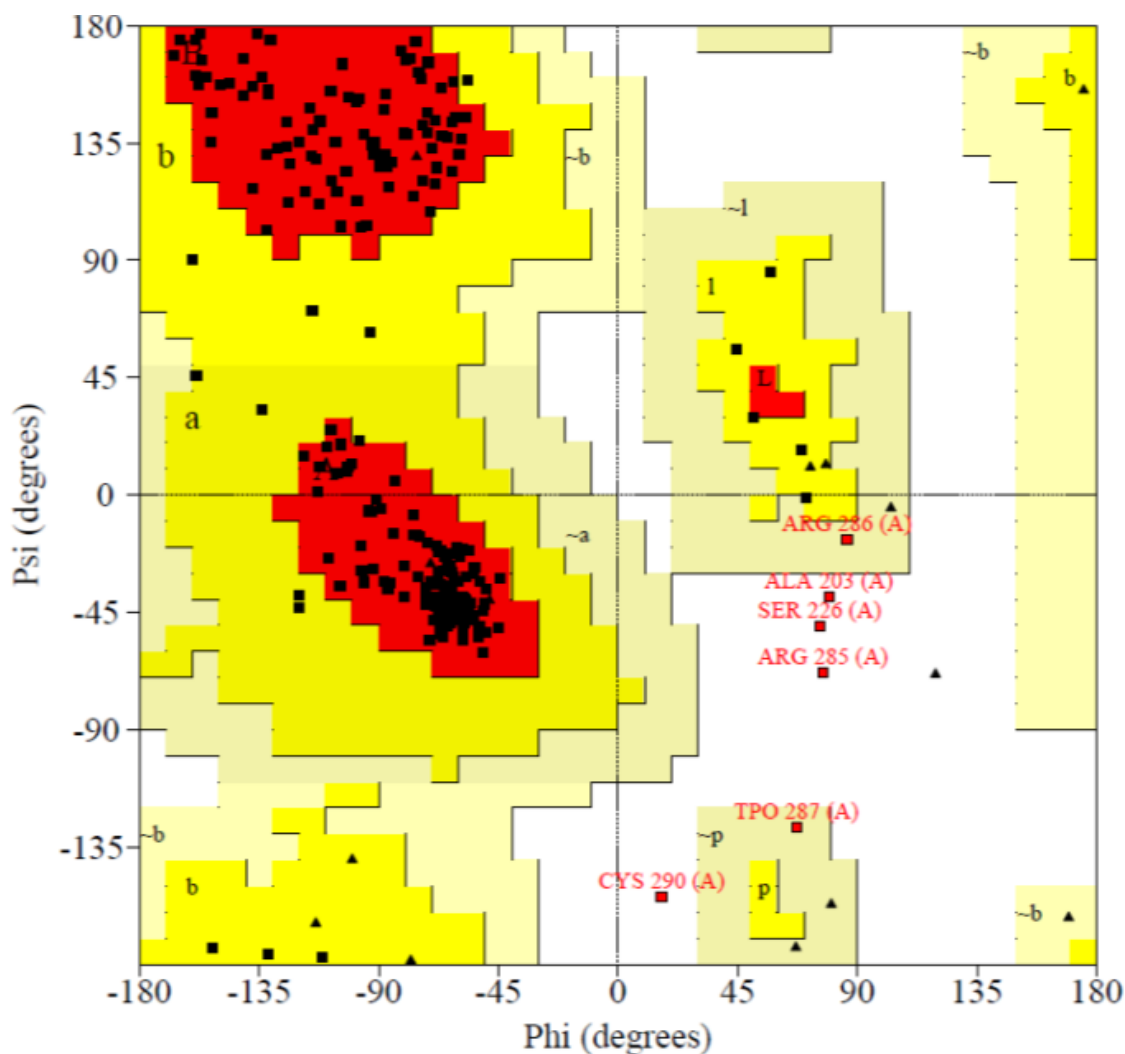


**Figure 2.6.** Ramachandran plot of 1OL7. The darkest areas (red) refer to the "core" regions, which represent the most favourable combinations of  $\phi/\psi$  values. The regions are labelled as follows: A - Core alpha, L - Core left-handed alpha, a - Allowed alpha, l - Allowed left-handed alpha, ~a - Generous alpha, ~l - Generous left-handed alpha, B - Core beta, b - Allowed beta, ~b - Generous beta, p - Allowed epsilon, ~p - Generous epsilon.

The general statistics of Ramachandran plot of 1OL7 is as follows:

		No. of residues	%-tage
Most favoured regions	[A,B,L]	193	82.5%*
Additional allowed regions	[a,b,l,p]	38	16.2%
Generously allowed regions	[~a,~b,~l,~p]	2	0.9%
Disallowed regions	[XX]	1	0.4%*
Non-glycine and non-proline residues		234	100.0%

End-residues (excl. Gly and Pro)	1
Glycine residues	15
Proline residues	14
	----
Total number of residues	264



**Figure 2.7.** Ramachandran plot of 4BN1. The darkest areas (red) refer to the "core" regions, which represent the most favourable combinations of  $\phi/\psi$  values.

The general statistics of Ramachandran plot of 4BN1 is as follows:

		No. of residues	%-tage
		----	----
Most favoured regions	[A,B,L]	211	90.6%
Additional allowed regions	[a,b,l,p]	16	6.9%
Generously allowed regions	[~a,~b,~l,~p]	2	0.9%
Disallowed regions	[XX]	4	1.7%*
		----	----
Non-glycine and non-proline residues		233	100.0%
End-residues (excl. Gly and Pro)		2	
Glycine residues		15	
Proline residues		13	
		----	
Total number of residues		263	

The overall stereochemical quality of 4BN1 is better than that of 1OL7 with 90.6% of residues in most favoured regions compared to that of 82.5% in 1OL7. However, local assessment of the activation loop showed that 4BN1 contained more outliers in the activation loop (R285, S286, T287 and C290) compared to 1OL7 (T287). The C290 residue is a critical requirement in this model as the biochemical data provided by Dr. Tsuchiya suggested a possible interaction between CoA and this residue. As a result, it is unlikely that 4BN1 can be used. Another parameter to check is the B-factors. The B-factors of the five regions of the catalytic domain of AURKA of 1OL7 and 4BN1, especially those of the residues in the activation loop were compared. 1OL7 provided similar B-factors to 4BN1 for the hinge region, flexible chain (green), back loop (purple), front loops (red and cyan). For the activation loop, which was of particular interest in this project, the B-factors for that of 1OL7 is mostly 60 – 80 Å<sup>2</sup> with the exception of T287 (80 – 100 Å<sup>2</sup>) and C290 (80 – 90 Å<sup>2</sup>). In comparison, the B-factors for that of 4BN1 is mostly 60 – 80 Å<sup>2</sup> with the exception of T287 and T288 and C290 (100 – 130 Å<sup>2</sup>). The full details of B-factor values for 1OL7 and 4BN1 are presented in Appendix 2.2. Overall the B-factors value for both pdb's showed the flexibility of the activation loop, particularly at T287, T288 and C290 residues. In 1OL7 the –SH sidechain of C290 was missing and the average B-factor of the main chain was 88 Å<sup>2</sup>. In 4BN1, even though the –SH sidechain was included (pointing outwards of the activation loop), its B-factor is 135 Å<sup>2</sup>. For C290 of 4BN1, the average B-factor of C290 was 108 Å<sup>2</sup> so the B-factor of –S atom of C290 was much higher than the average B-factor of C290. This presents a great uncertainty of the position of the –SH sidechain of C290. (Table 2.4) In summary, the B-factors of 1OL7 are lower than those of 4BN1 in activation loop, particularly in the phosphorylated region: T287, T288 and in the area of interest C290. It means that 1OL7 model provided similar or better accuracy of the positions of the residues in the activation loop. As a result, 1OL7 was chosen among AURKA and AURKB crystal structures and the sidechain of C290 was constructed using MOE programme. 1OL7 would also be used for modelling of CoA's conformation in AURKA and to test whether the modelling results obtained able to explain biochemical observations provided by Dr. Tsuchiya.

**Table 2.4 – Comparison of B-factor values of 1OL7 and 4BN1**

<b>B-factor value/ Å<sup>2</sup></b>	<b>1OL7</b>	<b>4BN1</b>
<b>Average B-factor of activation loop W277 – D294</b>	68.8	76.4
<b>Average B-factor of T287</b>	91	92
<b>Average B-factor of T288</b>	89	122
<b>Average B-factor of C290</b>	88	108
<b>B-factor of sidechain –S atom of C290</b>	n/a	135

In order to verify the quality of the structure of 1OL7 further, it was later used in docking programme to predict the conformations of CoA and its analogues in AURKA and in *in-silico* mutant experiment of the T217 residue of AURKA and to compare if the docking results can explain the biochemical findings. This was to confirm whether 1OL7 satisfied the “biological requirement” of a good model suggested by Kleywegt.

*Selection of protein structures for AURKB, Src, FGFR1, CDK2 and PKA to compare with AURKA*

The filtering of the protein structures followed similar process as that of the evaluation of Aur-A structure. The selected structures satisfied the crystallographic and statistical requirements such as resolution, R-factor and R-free as well as protein structural requirement (Ramachandran plot). They were in their phosphorylated/active form and contained a full activation loop or the most completed activation loop structure available. For AURKB, there are currently no human AURKB crystal structures so the best alternative was that of *Xenopus laevis*. It is of 68% sequence identical to human AURKB, sharing identical residues in the hinge region and sharing highly similar residues in the activation loop. The crystal structures of AURKB of *Xenopus laevis* available were limited (11 structures) but of good overall quality. The best one was pdb 4C2W<sup>111</sup> (Resolution 1.70 Å, R-factor 0.158, R-free 0.195, average B-factor of the hinge region E171 – E177 ~15 Å<sup>2</sup>, average B-factor of the activation loop W237 – D267 ~20 – 30 Å<sup>2</sup>). For Src, the selected structure was pdb 2H8H (Resolution 2.20 Å, R-factor 0.207, R-free 0.272, average B-factor of the hinge region E339 – S345 ~ 30 – 40 Å<sup>2</sup>, average B-factor of the activation loop L407 – T440 ~50 – 60 Å<sup>2</sup>).<sup>112</sup> For FGFR1, the selected structure was pdb 5AM6 (Resolution 1.96 Å, R-factor 0.215, R-free 0.251, average B-factor of the hinge region E562 – N568 ~ 35 Å<sup>2</sup>, average B-factor of the activation loop L644 – T678 ~65 Å<sup>2</sup>).<sup>113</sup> For CDK2, the selected structure was pdb 4I3Z (Resolution 2.05 Å, R-factor 0.199, R-free 0.228, average B-factor of the hinge region E81 – D86 ~ 30 Å<sup>2</sup>, average B-factor of the activation loop L148 – S181 ~35 – 45 Å<sup>2</sup>).<sup>114</sup> For PKA, the selected structure was pdb 2GU8 (Resolution 1.96 Å, R-factor 0.215, R-free 0.251, average B-factor of the hinge region E121 – E127 ~ 40 Å<sup>2</sup>, average B-factor of the activation loop F187 – N216 ~40 – 50 Å<sup>2</sup>).<sup>115</sup>

### 2.3.1.2. Structural comparison of AURKA, AURKB, Src, FGFR1, CDK2 and PKA

In the first instance, structural alignment of AURKA, AURKB, Src, FGFR1, CDK2 and PKA showed that the hinge region and the flexible chain where ATP bound, the back loop and front loop were highly conserved among the 6 kinases. For the activation loop, the similarity of the residues dropped significantly. Only AURKA, AURKB and PKA showed some similarity to each other. (Figure 2.8)



**Figure 2.8.** Structural comparison of the catalytic domain of AURKA, AURKB, Src, FGFR1, CDK2 and PKA. Box: DFG motif prior to the activation loop.

Examining the hinge region in detail, the sidechains of E211 and A213 of AURKA point downwards and away from the hinge region. It was known in many crystal structures of AURKA and ATP/ADP that the ligands were bound to the hinge region by hydrogen bonding with the backbone of E211 and A213. The equivalent residues to E211 and A213 of AURKA in AURKB, Src, FGFR1, CDK2 and PKA also have the sidechains pointing away from the catalytic pocket. In fact, no sidechains of the residues in the hinge region of AURKA (E211 – T217) pointed into the catalytic pocket. The sidechains of the equivalent residues to E217 of CDK2, PKA and AURKB (D86 of CDK2, E127 of PKA and E177 of AURKB) pointed into the ATP binding site. Those of FGFR1 and SRC pointed away. Preliminary biochemical findings of the IC<sub>50</sub> difference between CoA and dpCoA (4.4 and 17 μM respectively) suggested that there might be hydrogen bonding interaction between T217 and the 3' phosphate of CoA. It suggested that a T or S residue with –OH sidechain might be able to form hydrogen bonding with the 3' phosphate and account for the kinase sensitivity towards CoA. However, the structural alignment showed that although the T217 position was less conserved compared to other positions of the hinge region, it was difficult to conclude that there was hydrogen bonding between CoA and the residue at that position and such interaction was responsible for a kinase sensitivity towards CoA. The structural alignment showed that the –OH sidechain and the –NH mainchain of T217 pointed away from the catalytic region, making it unlikely that there would be any hydrogen bonding between T217 and 3' phosphate of CoA.

When examining the structural alignment of the activation loops of AURKA, AURKB, Src, FGFR1, CDK2 and PKA, the DFG motif prior to the activation loop was conserved among all investigated kinases but the activation loop residues varied. (Figure 2.9) The two kinases, AURKB and PKA, which shared the most residue similarity to AURKA in the activation loop, were examined in more details. For the activation loop of AURKA, only the sidechains of W277 and C290 were likely to point into the catalytic domain. For PKA, a number of sidechains pointed into the catalytic domain, including F187 (equivalent to W277 in AURKA), K189, T197 (phosphorylated site) and C199 (equivalent to C290 in AURKA). Among those residues, F187 and C199 were closest to the ATP binding site. For AURKB, W237 (equivalent for W277 in AURKA) and C250 (equivalent for C290 in AURKA) were also the closest residues to the ATP binding site with their sidechains pointing into the catalytic region. However, the exact positions of these residues (so as their distance from the ATP binding site) varied as the activation loop of the kinases adopted different conformation. None of the 11 crystal structures of AURKB (of *Xenopus laevis* species) available shared similar conformations of the activation loop with 1OL7. The AURKB kinase was in fully activated form and the conformations of its activation loop resembled that of 1OL5 (TPX2-bound activated AURKA) more. For PKA, the structural alignment showed that PKA also contained a C199 which was at the same position as C290 of AURKA and PKA was not CoA sensitive.



Further search on the different conformations of PKA activation loop out of 172 structures in MOE Protein Kinase Database revealed that none was reported to adopt similar conformations as that of AURKA in 1OL7. The activation loop of AURKA alone already showed varied conformations as presented in Figure 2.6. The conformation of the activation loop may be an important factor accounting for inhibitory selectivity of CoA towards AURKA.

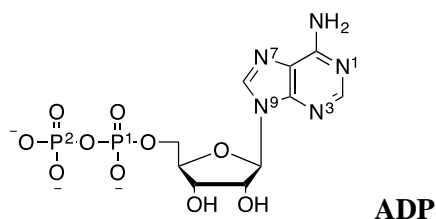
The observation of the flexible conformations of the activation loop also suggested that it might be very difficult to build an accurate homology model of the activation loop of AURKA or more precisely a more accurate model than which was obtained from X-ray crystallography. As mentioned, it is only possible to build a homology model if there exist homologous proteins which have similar conformations of activation loop of AURKA, then a model of the unknown activation loop can be constructed using amino acid alignment and the backbone structure of the known structures. The templates should share at least 40% homology with the targets and RMSD values between final positions of C $\alpha$  in the model and templates should be less than 1.5 Å. However, the activation loop appeared to be poorly conserved and is capable of adopting different conformations, which are unique for a particular kinase such as for AURKA, and essential for the activity of the kinase. This nature was not supportive to homology modelling.

*In summary,*

The hinge region was highly conserved in kinase structure so that the overall major effect that differentiated the fitting of CoA in AURKA catalytic domain over the catalytic domains of other kinases was likely to be the steric effect around T217 region. This steric effect around T217 region likely defines the pocket around the hinge region for CoA to fit in.

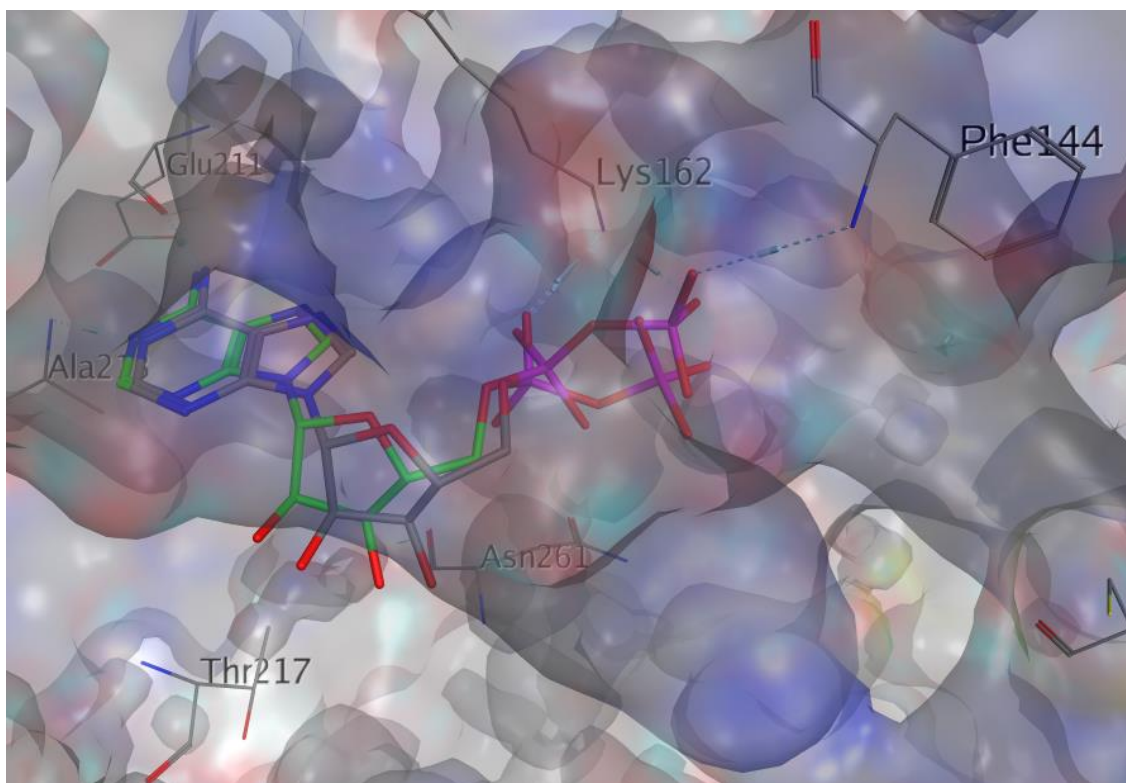
The flexible and various conformations of the activation loop might be an important factor for CoA selectivity. AURKA kinase seemed to adopt a unique conformation of the activation loop that allowed interactions with CoA.

### 2.3.2. Docking results and discussion of the conformation of CoA in AURKA



The validation of CHEMPLP showed that CHEMPLP correctly predicted the binding conformation of the adenosine group of ADP (the original ligand) as that in the X-ray crystal structure. ADP bound at the hinge region of the ATP binding site, with the –NH<sub>2</sub> of adenine ring forming H-bond with –CO backbone of E211. (Figure 2.9) For the flexible ribose and pyrophosphate groups, CHEMPLP predicted the binding position similar to that in the original

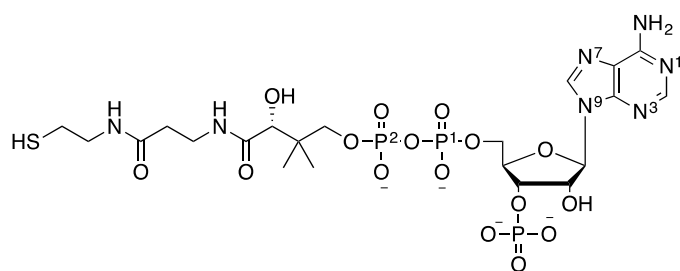
crystal structure. The ribose was predicted not having any hydrogen bonding interactions and this is in agreement with the binding shown in the original structure. The pyrophosphate group was predicted to form hydrogen bonding between P<sup>1</sup>=O and –NH<sub>2</sub> sidechain of K162. This is in agreement with the original crystal structure. CHEMPLP predicted that the pyrophosphate group might take up another hydrogen bonding between P<sup>2</sup>=O and –NH backbone of F144. This was not indicated in the original crystal structure. Overall, CHEMPLP was able to predict accurately the binding of rigid structural part of ADP and was able to predict the position and binding of the flexible structural part of ADP to a reasonable extent.



**Figure 2.9.** Re-evaluation of 1OL7. ADP docking result. a) ADP crystallised structure (C atoms in grey), ADP solution generated by CHEMPLP (C atoms in green). H-bonds (dashed lines).

CHEMPLP was also able to predict the conformations of other co-crystallised molecules of other original protein structures (in this project, AURKA) of various expression systems; for example, molecule AK8 in 3EFW (expression system: *Spodoptera frugiperda*),<sup>99</sup> YPH in 4JBP (*Homo sapiens*),<sup>100</sup> and ZZL in 2WTV (*Escherichia coli*)<sup>29</sup>. (Appendix 2.3) As CHEMPLP was able to predict the conformation of ADP in 1OL7 and the conformations of other binding ligands in other protein structures, it was then used to dock CoA and its analogues in the two models of 1OL7, original one with missing sidechain of C290 and modified one with the sidechain of C290 rebuilt by MOE. The aim was that the obtained conformation result of CoA would be able to explain the biochemical findings that CoA exhibited inhibition against AURKA in the absence of TPX2 and that IC<sub>50</sub> (CoA) was increased (4.4 μM to 47 μM) in the presence of reducing agent DTT.

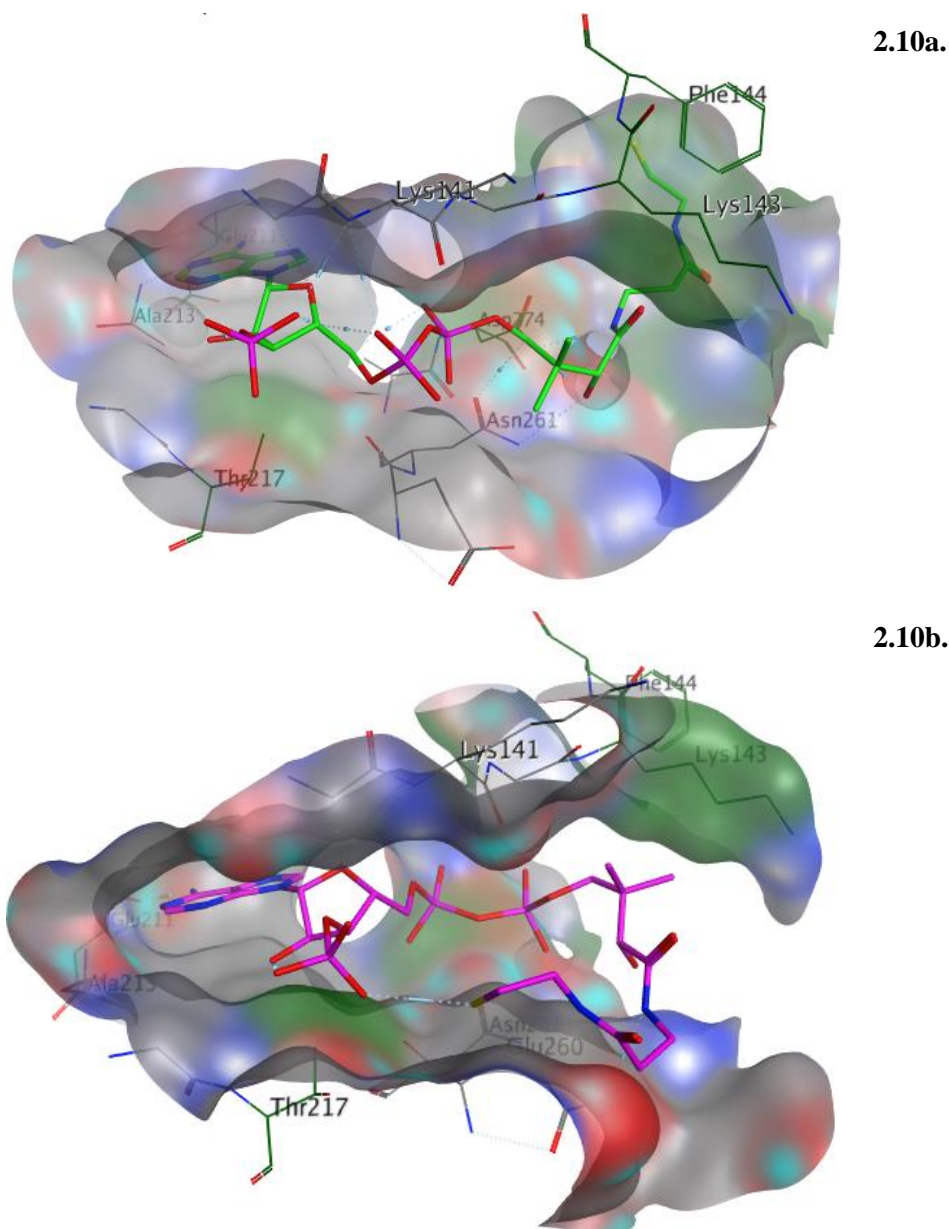
### 2.3.2.1. Noncovalent docking results of CoA and its analogue desulfo-CoA in AURKA



#### *Original crystal structure IOL7 without C290 sidechain correction*

The result showed that the adenosine group of CoA bound to the kinase hinge region in a similar manner to ATP/ADP. Out of 10 docking solutions, there were 5 solutions where CoA sat in the binding site. The best 3 binding poses from CoA, determined by the largest docking scores, were saved and examined.

The docking solutions can roughly fall into two binding modes of CoA in the catalytic domain of AURKA: binding mode **I** and **II** as illustrated in Figure 2.10. Out of the 5 docking solutions, four exhibited binding mode **I** and one exhibited binding mode **II** (defined in next sentences). In both binding modes, CoA bound to the hinge region through H-bonding in a similar fashion to ADP. The ribose sugar with 3'phosphate sat at the region below G140 and Lys141 of flexible chain and above T217 region. Binding mode **I** suggested that the 3'phosphate might form H-bond with –NH backbone of K141 whilst binding mode **II** suggested that it might form an internal H-bond with –SH of CoA. In binding mode **I** (higher fitness score: 87.2, Figure 2.8a), the pantetheine tail sat in the back loop. Its direction was secured by H-bonds of –P<sup>2</sup>=O with –NH backbone of K141 and –OH with –NH of N261 and –OH and –NH with –COOH sidechain D271. In binding mode **II** (lower fitness score: 79.3, Figure 2.8b), similar interactions between CoA and hinge were observed. The ribose sugar with 3'phosphate sat in similar location as in binding mode **I** but the 3'phosphate curved down, forming internal H-bonds with –OH of the ribose sugar and –SH of the pantetheine tail instead. The pantetheine tail curved to the front surface in order for the thiol group to form the H-bond with the 3'phosphate.

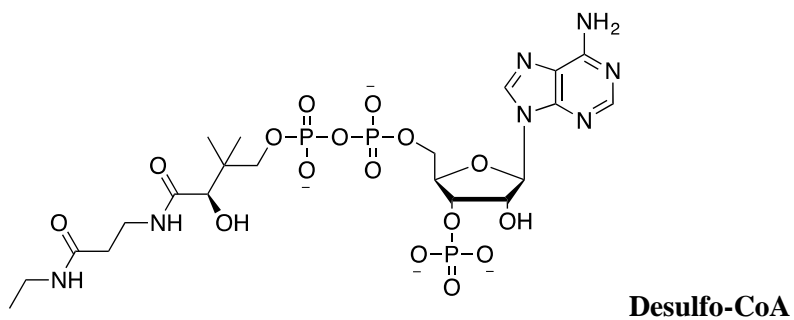


**Figure 2.10.** Two binding modes of CoA suggested by GOLD. a) Binding mode **I** (C atoms in green). b) Binding mode **II** (C atoms in orange). H-bonds (dashed lines).

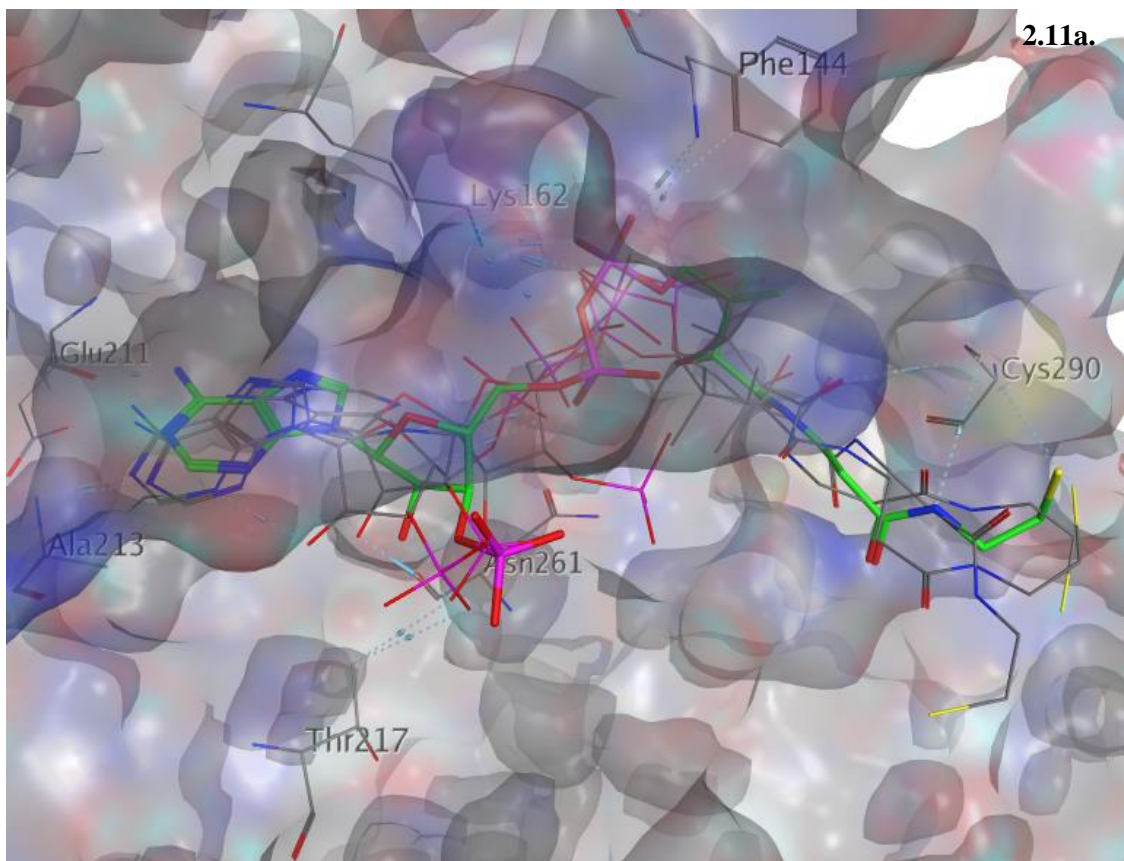
*1OL7 with rebuilt C290 sidechain*

The non-covalent docking solutions of CoA suggested the flexibility of the pantetheine tail and a possibility for it to reach to the activation loop.

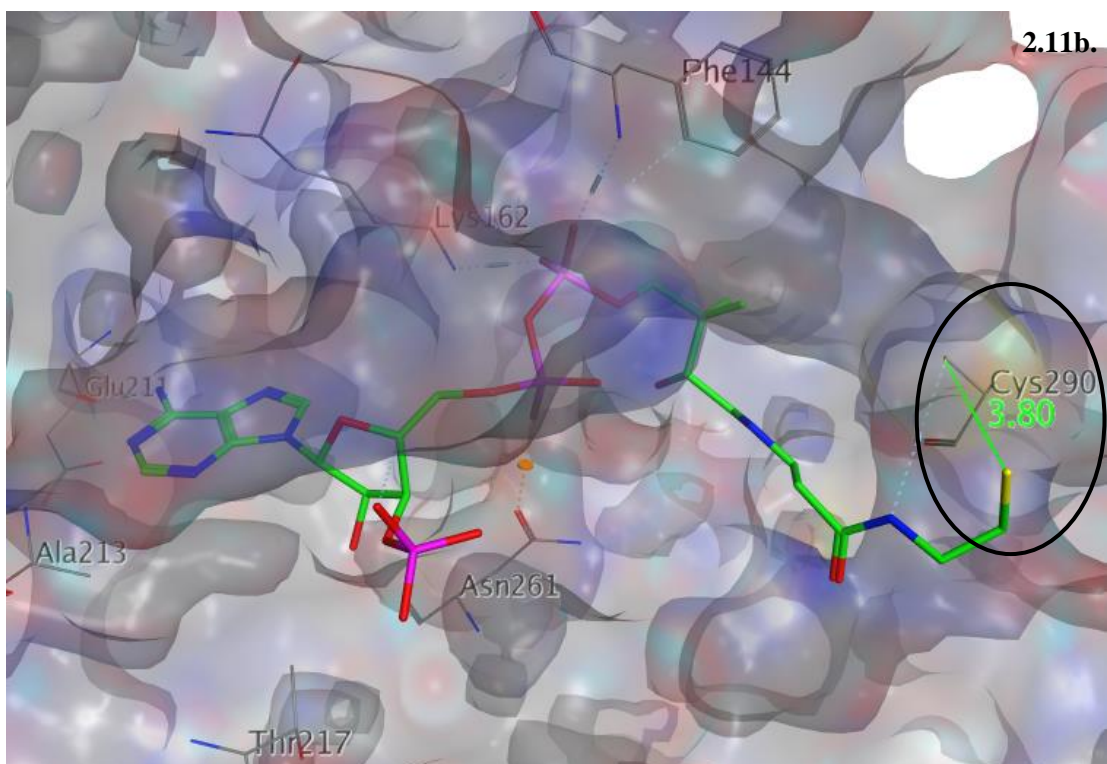
Using this 1OL7 structure, GOLD docking was carried out with ADP, CoA and desulfo-CoA (CoA molecule without –SH group) to test any interactions between –SH of CoA and C290.



Out of 10 docking solutions, there were 5 solutions where the conformations of CoA extended to the C290 residue. (Figure 2.11a) In other words, 50% of the best docked poses showed that there would be a potential for interaction between CoA (-SH) and C290 of AURKA (non-covalent or covalent i.e. di-S bond). The best binding poses from CoA, determined by the largest docking scores, suggested possible disulfide linkages with C290 (Figure 2.11b) based on the close distance between the two sulfur atoms of CoA and C290 of 3.8 Å.

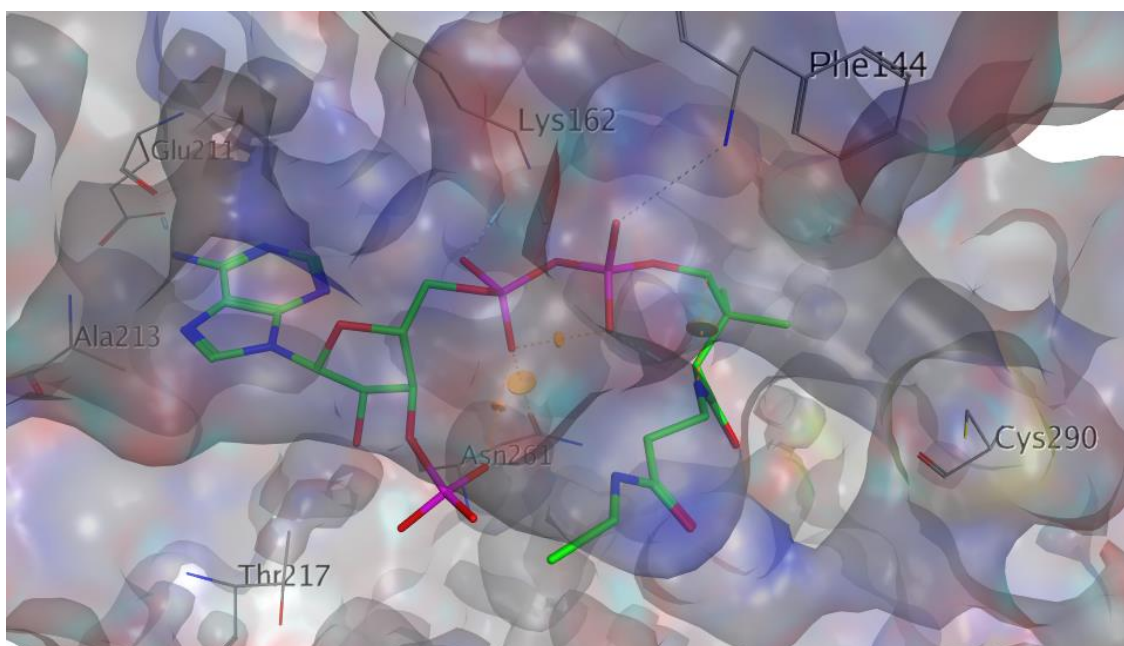






**Figure 2.11.** Binding poses of CoA suggested by GOLD showing position of –SH around C290 in AURKA ATP binding pocket. a) Highest scored docking pose (85) (C atoms in green). Other scored docking poses (C atoms in grey). b) Highest scored docking pose of CoA generated by GOLD. H-bonds: dashed lines. Clashes: orange rings.

For desulfo-CoA, however, the best docking pose was curved to the front. Forty percent of the best docking poses had the tail around the C290 area (shown below), similar to the best pose of CoA. However, not having the –SH, there was no possibility for desulfo-CoA to form di-S bonding with C290 of AURKA.



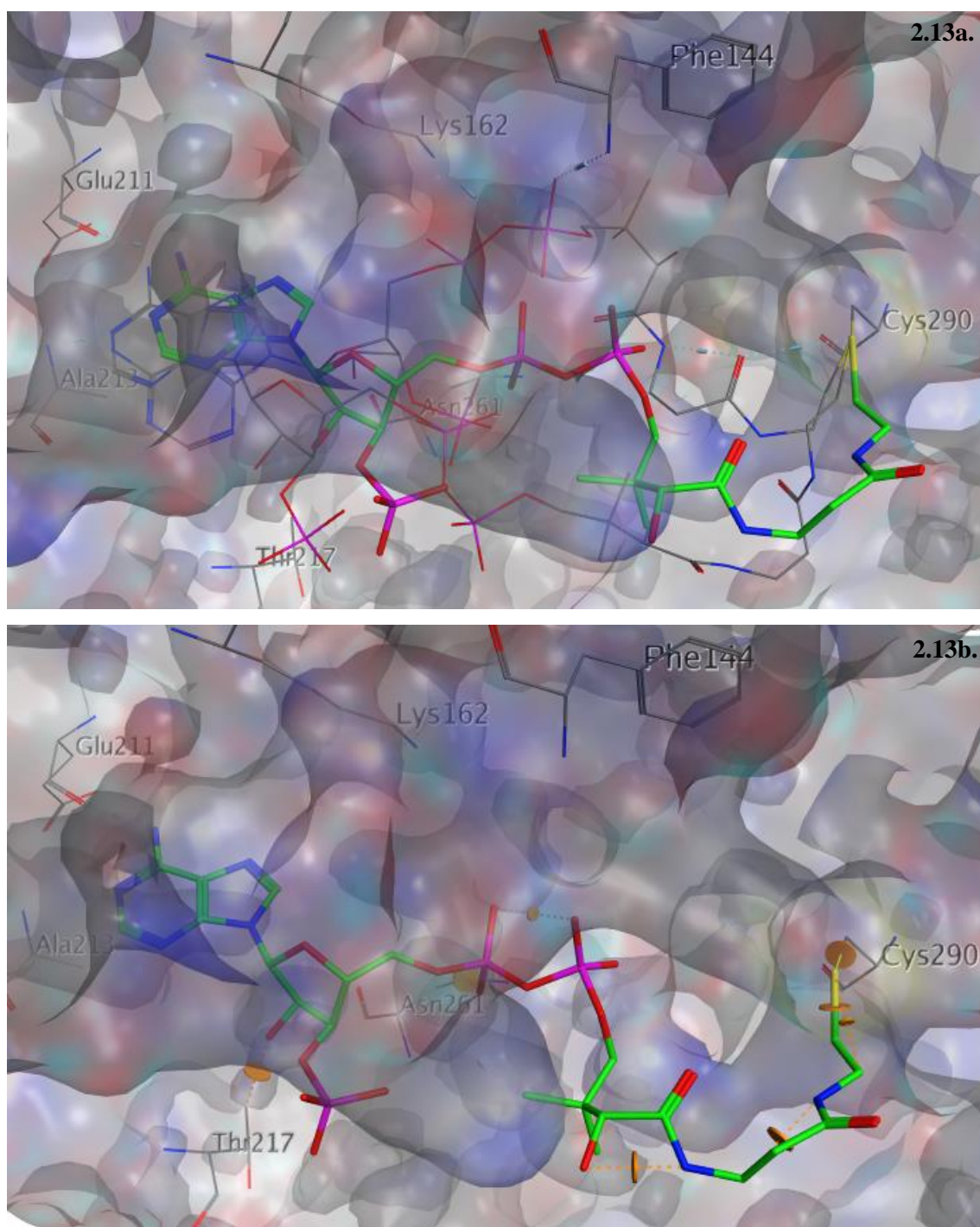
**Figure 2.12.** Best binding pose of desulfo-CoA suggested by GOLD showing the possible position of end-CH<sub>3</sub> in AURKA ATP binding pocket. Highest scored docking pose (73.4) (C atoms in green). H-bonds (dashed lines). Clashes (orange rings).

### 2.3.2.2. Covalent docking results of CoA and its analogue desulfo-CoA in AURKA

This experiment was to test if a covalent bond was formed first whether CoA would fit inside the ATP binding pocket. It was hypothesised that when a covalent bond was formed first, CoA would NOT fit inside the ATP binding pocket of AURKA. The docking was done on 1OL7 with rebuilt C290 model.

#### *Docking results*

Three out of 10 solutions and 3 out of 5 highest scored binding mode that suggested CoA might penetrate into the ATP binding pocket, particularly the hinge region. (Figure 2.13)



**Figure 2.13.** Binding poses of CoA inside AURKA ATP binding site suggested by GOLD when a di-S covalent bond is set between  $-SH$  (CoA) and C290. a) Three scored docking poses of CoA that fit in the hinge region of the ATP binding pocket: best fit (*cf.* binding of ADP): docking score of 88.6 (C atoms in green). Other two scored poses (scored 91.8 and 87.4) (C atoms in grey). b) Best fit docking pose (C atoms in green). H-bonds (dashed lines). Clashes (orange rings).

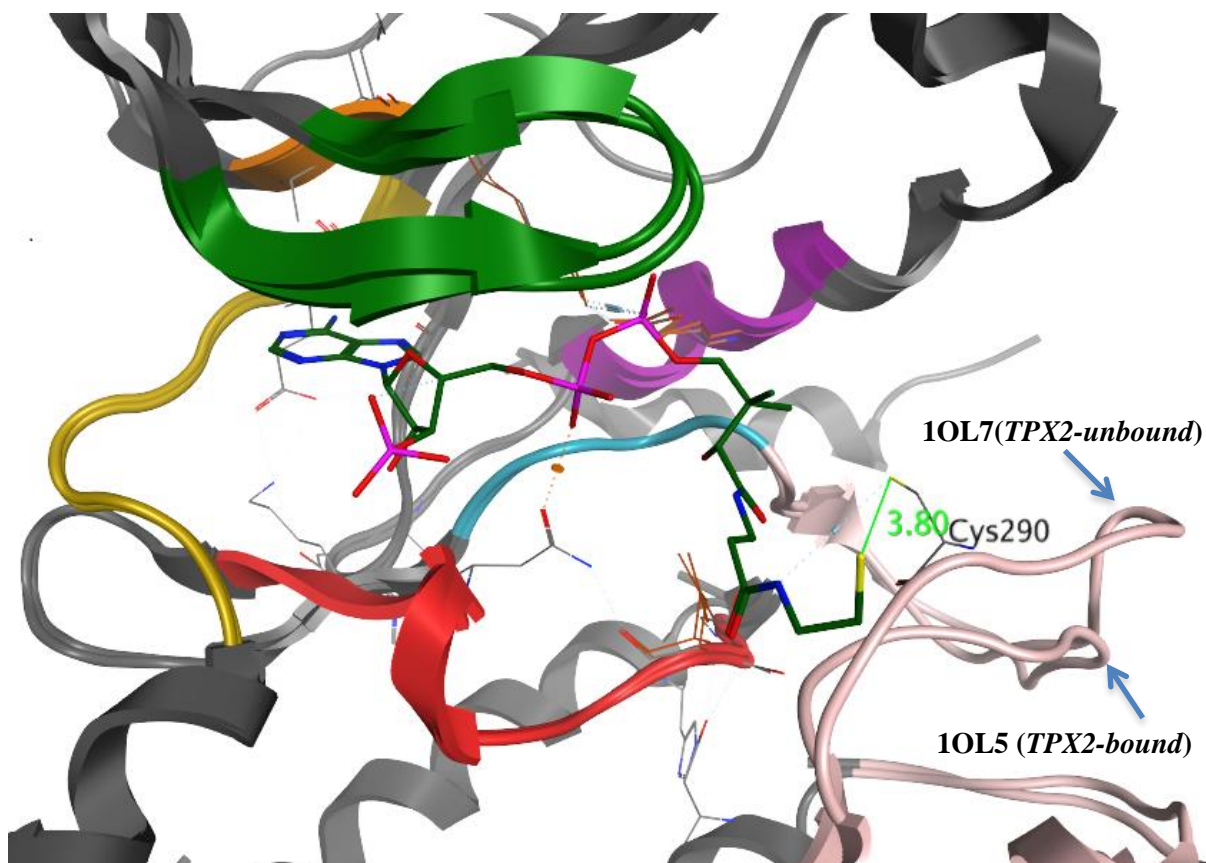
Out of the three solutions, only one showed a conformation at the adenosine head to be close to that of ADP in the hinge region. However, in order to accommodate such conformation when a di-S bond is formed between CoA and C290 first, the rest of the molecule curved out of the



pocket, creating bond tensions/clashes. In conclusion, in most cases, the adenine head of CoA did not manage to penetrate in ATP hinge region. (Figure 2.13b)

### ***2.3.2.3. Discussion of docking results***

In summary, the non-covalent docking results suggested that CoA was able to fit in the substrate binding site of AURKA, likely in an extended conformation with the adenosine head located in the hinge region through hydrogen bonding with E211 and A213 and the pantetheine tail located in the activation loop. There would be some interactions between the –SH of CoA and the sidechain –SH of C290 of the activation loop due to the close distance (3.8 Å) of these two when CoA was in the extended form. (Figure 2.11) This interaction with the activation loop may be important for the selectivity of CoA towards AURKA as later structural comparison of the activation loop of Aurora kinases to other kinases in the screening panel showed that the activation loop of Aurora kinases was significantly different from others and C290 residue was not been seen in equivalent conformation in any other kinases. This suggested conformation of CoA with possible interactions between –SH of CoA and C290 residue was able to explain the increase in  $IC_{50}$  (CoA) from 4.4  $\mu$ M to 47  $\mu$ M in the presence of DTT (a reducing agent to reduce disulphide bond). Furthermore, this conformation was also able to correlate with the inactivity of CoA in the presence of TPX2. In the TPX2-bound state, the activation loop of AURKA was pulled down as seen in crystal structure 1OL5 hence there would be no more interaction between –SH of CoA and C290. (Figure 2.14)



**Figure 2.14.** Comparison of the possible interactions between CoA and the activation loop of AURKA in the conformations of the TPX2-unbound state (1OL7) and TPX2-bound state (1OL5).

Docking results also showed that when the  $-SH$  of C290 was removed (no interaction with C290), CoA adopted different conformations (suggested as binding mode I or II, Figure 2.10) and no longer interacted with the activation loop. These conformations might not be able to sustain CoA in the substrate binding site of AURKA. Moreover, using this extended conformation of CoA in covalent docking, the docking results suggested that it was unlikely that a di-S bond would form between  $-SH$  of CoA and C290 prior to the binding of adenosine head of CoA inside the hinge region. This is in agreement with the biochemical conclusion (finding (3), Section 1.4) where CoA was postulated to bind to the hinge region first before rearranging itself to form interactions at the activation loop.

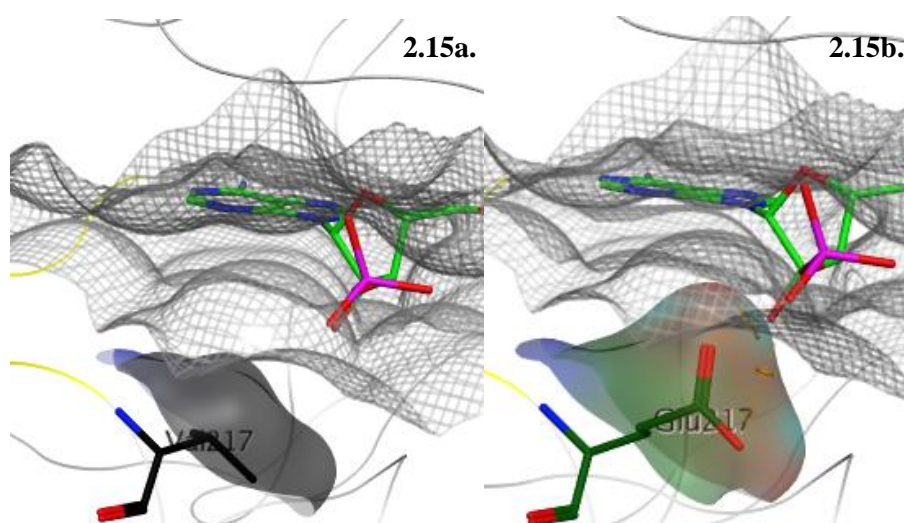
As this model was able to explain the biochemical findings, it was then used in the *in-silico* mutant experiment of T217 residue to investigate any possible interactions between CoA and T217 residue.

#### **2.3.2.4. *In silico* mutant experiment on AURKA T217**

The *in-silico* mutant experiment was done on 1OL7 structure with corrected C290 in MOE (using protein builder tool) and the preferred CoA's mode of binding in this case (Figure 2.11b)

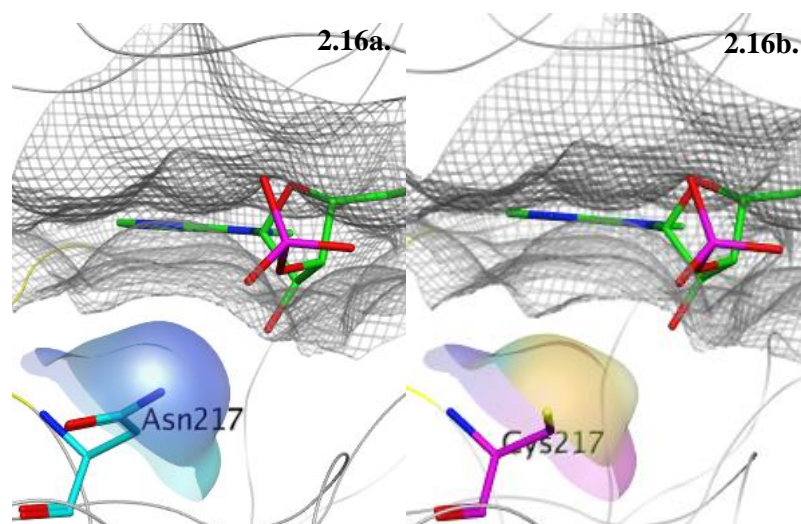
was used. The selected residue, T217 was mutated to T217N, T217V, T217E and T217C using protein builder tool. The possible interactions/clashes between those mutants and CoA were judged in electrostatic and steric terms.

The results from the *in-silico* mutant experiment suggested that at the T217 position, steric effects likely accounted for the difference in IC<sub>50</sub> between CoA and dpCoA instead of the hydrogen bonding. A supplementary check on the possible interactions between the proposed extended conformation of CoA and some T217 mutants was carried out. It was shown that the 3' phosphate of the proposed extended conformation of CoA did not show to form any hydrogen bonding with T217. When comparing AURKA T217 mutants: T217N, T217E, T217C and T217V with the wild type AURKA (1OL7), it appeared that T217V and T217E mutants had the sidechain pointing inwards the binding pocket, hence creating steric hindrance against CoA fitting. The clashes were particularly clear in the case of T217E mutant. (Figure 2.15b)



**Figure 2.15.** Surface map between AURKA T217V and T217E mutants and CoA. a) AURKA T217V (C atoms in black). b) AURKA T217E (C atoms in dark green). Part of CoA preferred binding mode (C atoms in green). Surface interaction of CoA (mesh).

For T217C and T217N mutants, however, the sidechains of T217C and T217N protrude outwards and away from the binding pocket, resulting in no steric clashes. As a result, although the length of the sidechain of N residue is similar to E, it would not obstruct CoA fitting. (Figure 2.16)



**Figure 2.16.** Surface map between AURKA T217N and T217C mutants and CoA. a) AURKA T217N (C atoms in cyan). b) AURKA T217C (C atoms in pink). Part of CoA preferred binding mode (C atoms in green). Surface interaction of CoA (mesh).

### 2.3.2.5. Overall conclusions

The non-covalent and covalent docking experiments and *in-silico* mutant experiment confirmed that the IOL7 model was able to explain the biochemical findings observed by Dr. Tsuchiya hence it was used further to design compounds with the aim to verify the CoA's conformation in AURKA and to produce an AURKA selective inhibitor. The results also suggested that:

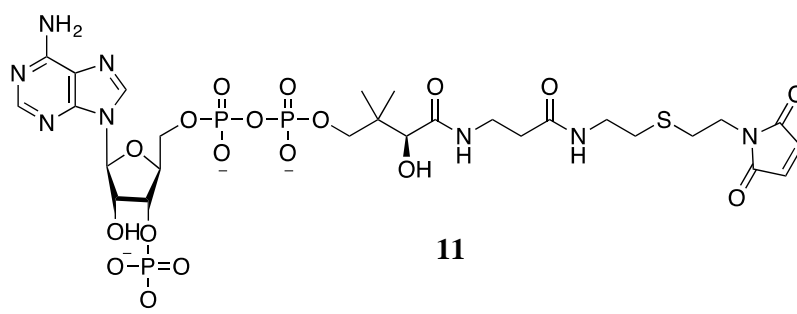
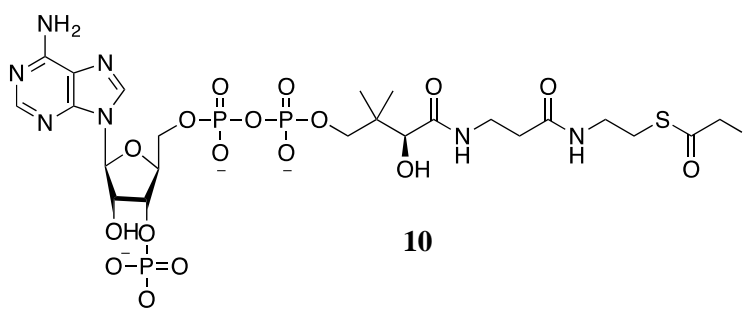
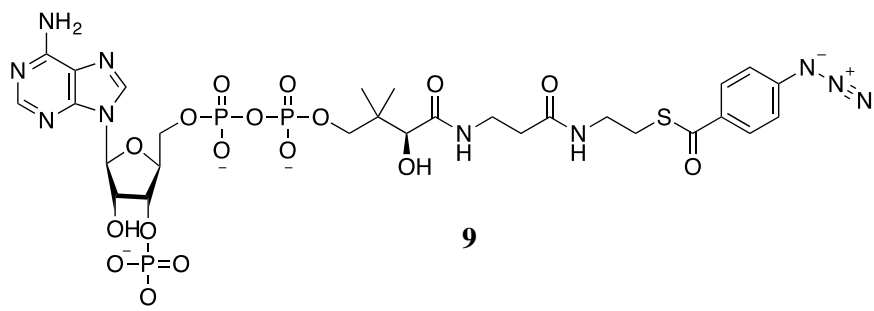
- (i) CoA was able to fit in the ATP binding pocket of AURKA in an extended conformation with the end –SH in close proximity to C290 of the activation loop. The length of CoA seems to be optimal for binding to AURKA ATP active site. Longer length seems to deplete the chance of contact with C290 of the activation loop.
- (ii) It was unlikely for CoA to form a di-S covalent bond first then rearrange itself to fit in the ATP active site of AURKA. Such procedure created too much strain to be compensated for by the formation of a covalent bond.

### 2.3.2.6. Further suggestion of CoA analogues to investigate the conformation of CoA in AURKA

Affinity probes, which are CoA analogues, were designed in order to further examine:

- (i) The upper limit of the length of a compound that would fit in AURKA catalytic domain.
- (ii) The possible location of CoA in AURKA without a crystal structure.
- (iii) The possible covalent bonding that can be generated around AURKA catalytic domain.

Three common (chosen) examples: -azido benzoic acid (a photoreactive crosslinker), -iodoacetamide and -maleimide analogues of CoA, were suggested for docking. Both non-covalent and covalent docking were carried out for the three analogues.

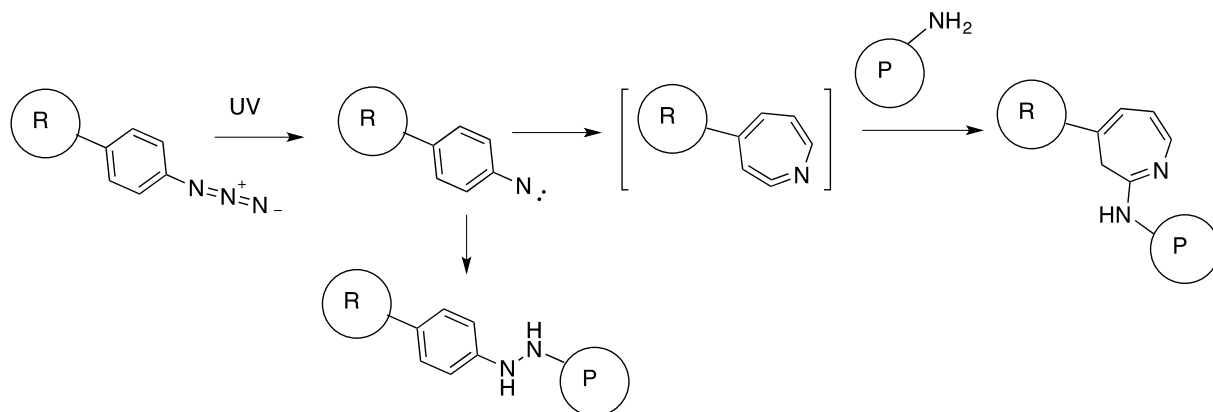


### Backgrounds <sup>116</sup>

Crosslinking is the process of chemically joining two or more molecules by a covalent bond. Crosslinking reagents (or crosslinkers) are molecules that contain two or more reactive ends capable of chemically attaching to specific functional groups (primary amines, thiols, etc.) on proteins or other molecules.

In this particular context, crosslinking is used for **capturing** and identify unknown protein interactors or interaction domains. The thiol group, -SH of the sidechain of C290 is the target site for crosslinking in this case. Examples of popular crosslinker reactive groups in this case are maleimide and haloacetyl (-bromo or -iodo) (**10** and **11** analogues). It is also of interest to investigate other possible interactions between CoA analogues and the catalytic domain of AURKA. As a result, a photo-reactive crosslinker, azidobenzoic acid **9** <sup>44b</sup> analogue, was taken into consideration.

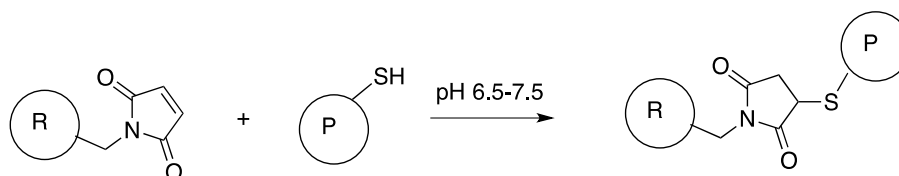
### *Aryl azide (Azidobenzoic acid) reaction chemistry*



**Figure 2.17.** Photochemical conjugation between aryl azide reagent and primary amine of protein residues.

Under exposure to UV light (250 – 350 nm), an aryl azide forms a nitrene group that can initiate addition reactions with double bonds or insertion into C-H and N-H sites or can undergo ring expansion to react with a nucleophile (e.g. primary amine). In the case where there is an exposed primary amine (from the side-chains of protein residues), the ring expansion reaction is favoured compared to other reactions.

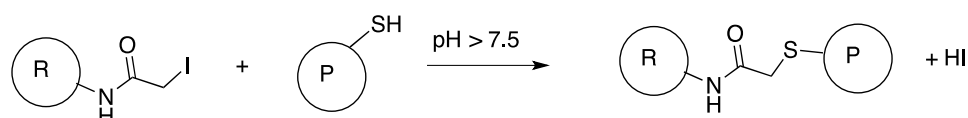
### *Maleimide reaction chemistry*



**Figure 2.18.** Thioether linkage reaction between a maleimide-containing reagent and a thiol group of a protein.

In a pH medium of 6.5 – 7.5, the maleimide group of the reagent reacts specifically with thiol groups of a protein. This results in the formation of a stable thioether linkage that is irreversible. In more alkaline conditions (pH >8.5), the reaction favors primary amines and also increases the rate of hydrolysis of the maleimide group to a non-reactive maleamic acid. Maleimides do not react with tyrosines, histidines or methionines.

### *Haloacetyl (-iodo) reaction chemistry*



**Figure 2.19.** Conjugation of iodoacetyl crosslinker and a thiol group of a protein.

The most commonly used haloacetyl crosslinkers contain an iodoacetyl or a bromoacetyl group. At physiologic pH (>7.5) haloacetyls react with thiol groups, resulting in a stable thioether

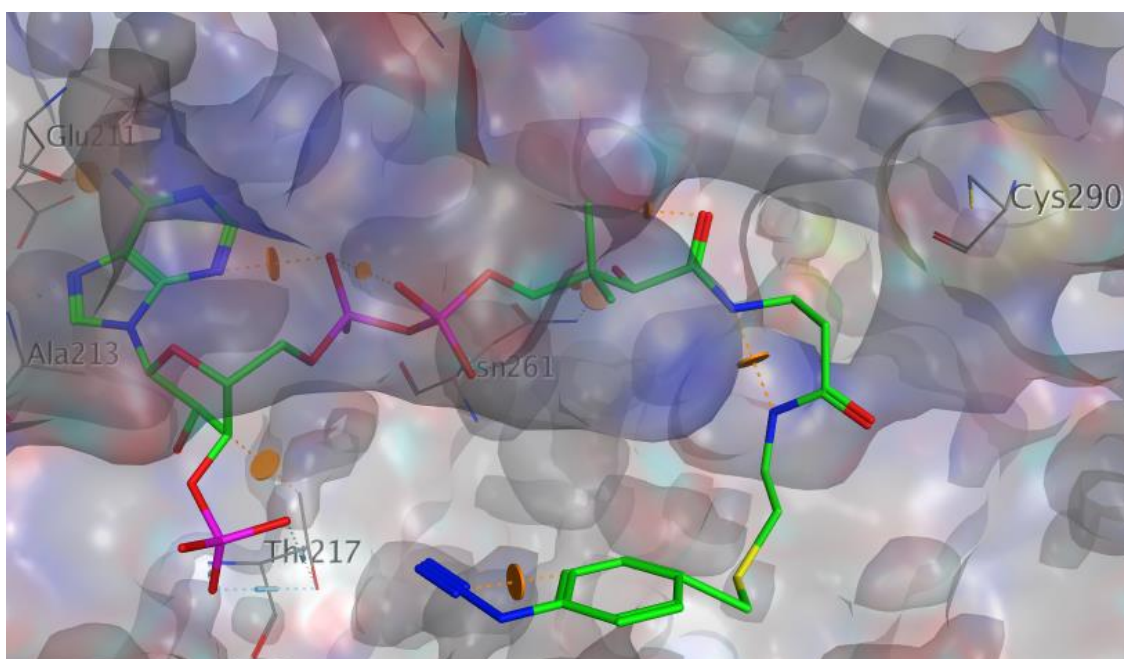


linkage. The iodoacetyl group is able to react with other amino acids in the absence of free thiols, or if iodoacetyl group is used in large excess. Histidyl sidechains and amino groups react in the unprotonated form with iodoacetyl groups above pH 5 and pH 7, respectively. Iodoacetyl reactions and preparations are performed in the dark in order to limit free iodine generation, which has the potential to react with tyrosine, histidine and tryptophan residues.

#### *Non-covalent docking docking results*

##### **Azidobenzoic acid analogue, 9**

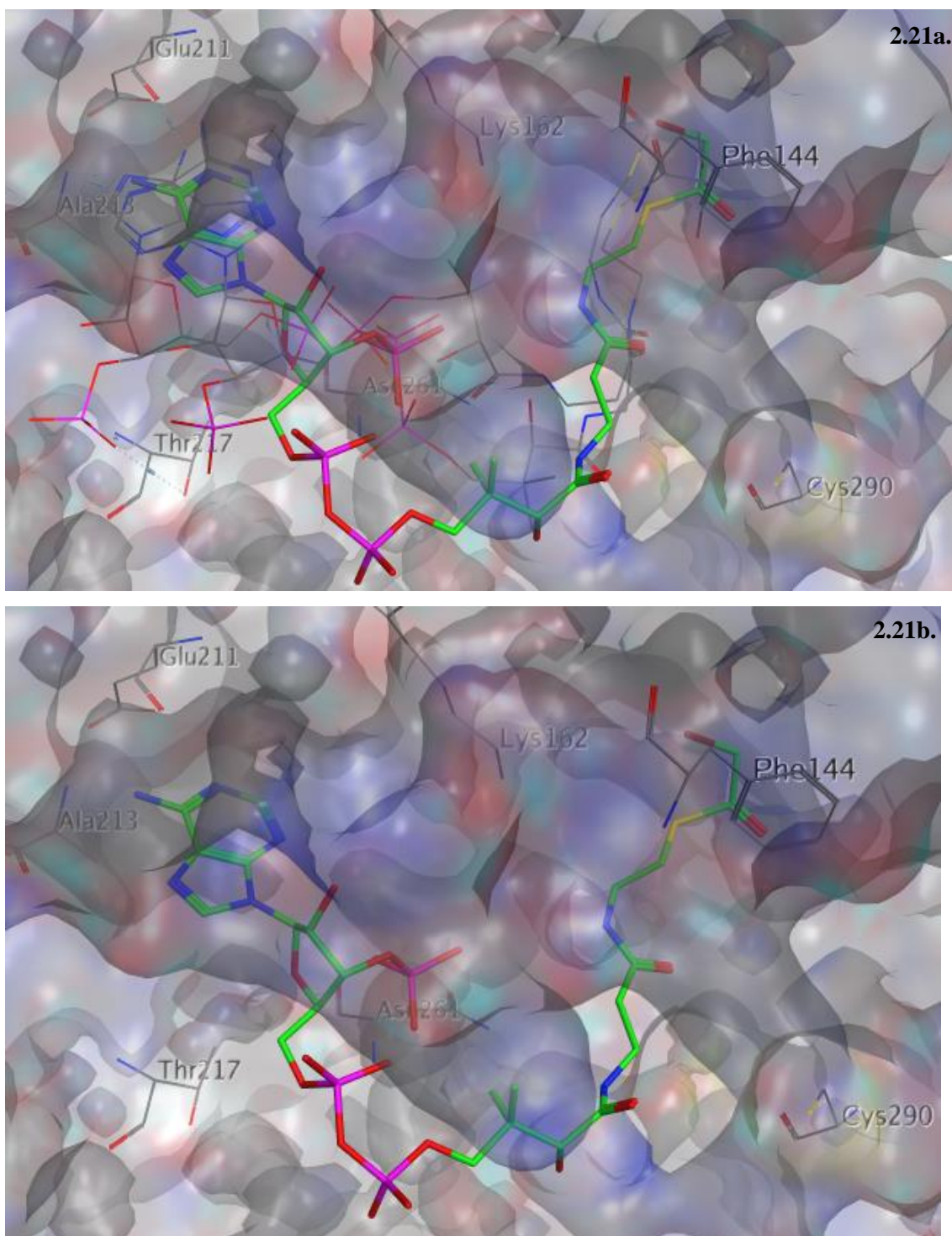
All 10 solutions suggested that analogue **9** could not fit in the binding pocket in the same orientation as CoA. Only 2 out of 10 scored docking poses suggested the possibility of the adenine head of **9** to fit in the hinge region. Even in that case, the length forced the molecule to curve around and out of the pocket. The end of the molecule containing the phenyl azide reactive group was not in close proximity to C290 in any scored docking poses.



**Figure 2.20.** Highest scored docking pose (82.9) of CoA's azidobenzoic acid analogue, **9** suggested by GOLD (C atoms in green) in non-covalent bonding setting. H-bonds (dashed lines). Clashes (orange rings).

##### **Iodoacetimide analogue, 10**

There is 60% chance of finding iodoacetimide analogue, **10** with the adenine head group in the ATP hinge region and three out of four best scored docking poses, including the highest scored docking pose, suggested that the tail, due to its length, would curve into the back of the ATP binding site and not be in close contact with C290. (Figure 2.21)



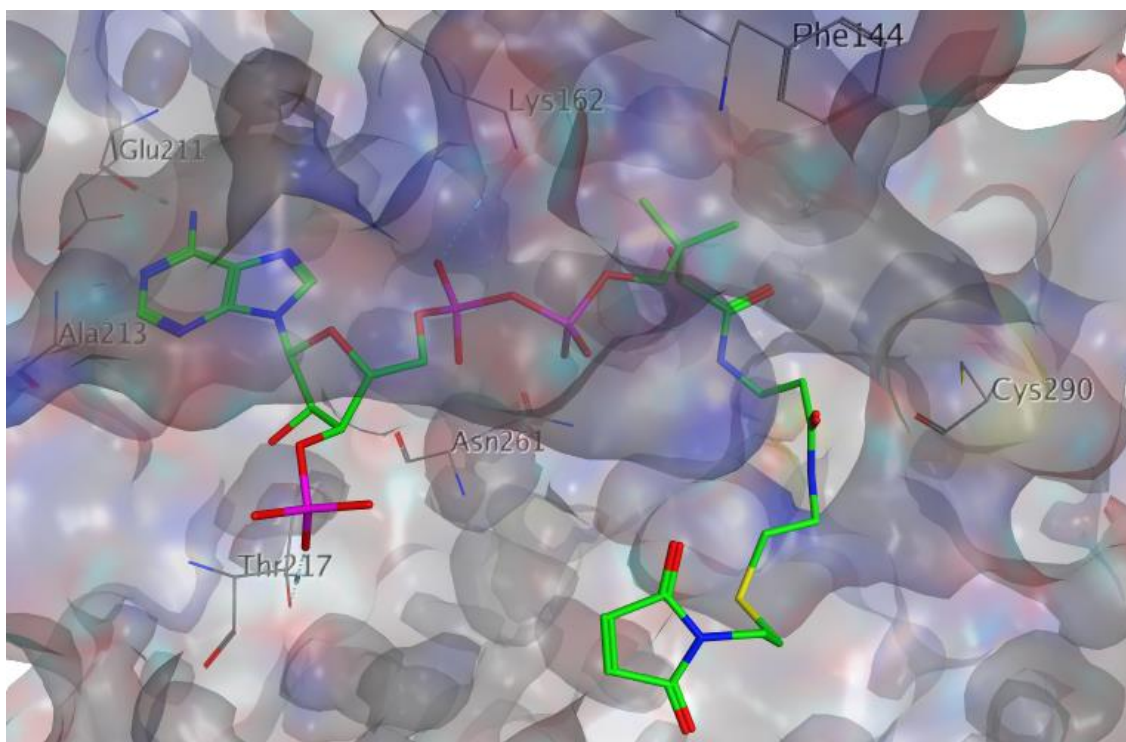
**Figure 2.21.** Binding poses of **10** inside AURKA ATP binding site suggested by GOLD in non-covalent bonding setting. a) Three best binding poses (out of 10) of CoA's iodoacetimide analogue, **10** suggested by GOLD. Highest scored docking pose (73.7) (C atoms in green). Other scored docking poses (C atoms in grey). b) Highest scored docking pose. H-bonds (dashed lines). Clashes (orange rings).

#### Maleimide analogue, **11**

There is only 1 out of 10 solutions (the highest scored docking pose) suggesting the binding of maleimide analogue, **11** to AURKA ATP binding pocket. However, due to the length, the



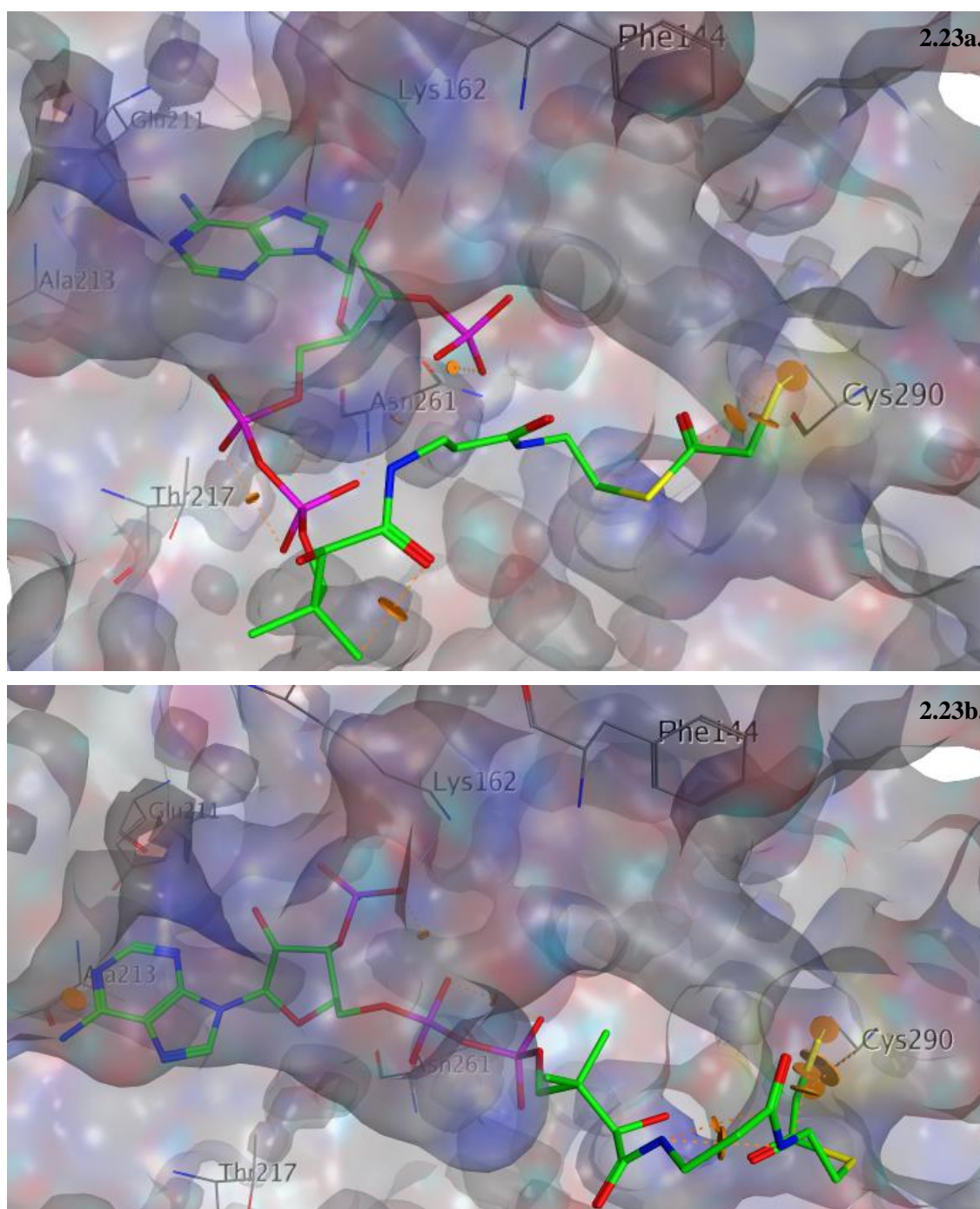
maleimide tail also had to curve to the front, out of the pocket and not in close proximity to C290. (Figure 2.22)



**Figure 2.22.** Single valid scored docking pose (78.6) of CoA's maleimide analogue, **11** suggested by GOLD (C atoms in green) in non-covalent bonding setting. H-bonds (dashed lines). Clashes (orange rings).

#### *Covalent docking results*

The covalent docking results showed that the analogues have to come out of the binding pocket completely in order to accommodate the set covalent bond at C290. Even for iodoacetimide analogue, **10**, the shortest among the three, none of the scored docking poses showed that it may be able to fit in the hinge region. Figure 2.23 showed the two highest scored docking poses (scored 105.8 and 99.0 respectively in Figure 2.23a and 2.23b) of A2 in this setting which are also the closest to the hinge region compared to the rest. However, those were not able to fit in the hinge region in the manner seen in ADP.



**Figure 2.23.** Two highest scored binding poses (out of 10) of CoA's iodoacetimide analogue, **10** suggested by GOLD when a di-S covalent bond was set between –SH (A2) and C290. a) Highest scored docking pose (105.8). b) Second highest scored docking pose (99.0) (C atoms in green).

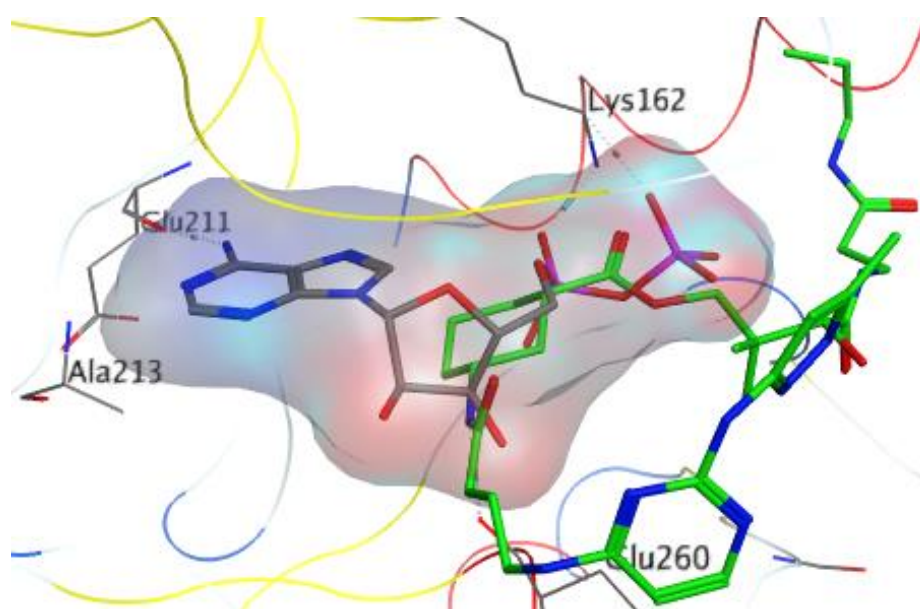
*In summary,*

Out of the 3 examined proposed affinity analogues of CoA, it is shown by docking that the iodoacetimide analogue, **10** and azido benzoic acid analogue, **9** may be able to bind to ATP active site and hence suggested for synthesis. (The synthesis was carried out by my colleague, Fiona Bellany.)

## 2.4. *Ab initio* design of a new AURKA selective inhibitor

### 2.4.1. Re-examination of preliminary docking results in the 1<sup>st</sup> generation of compounds

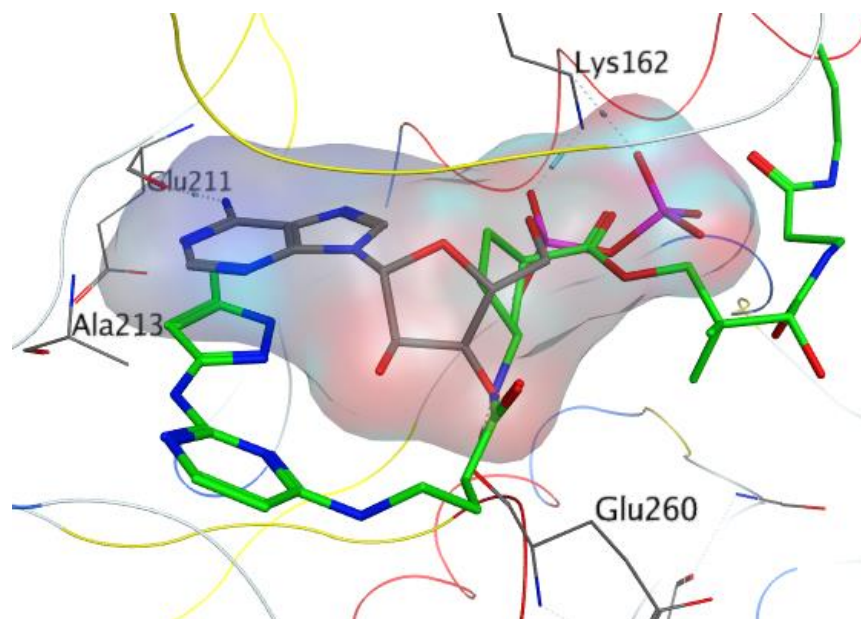
There were three critical problems with the initial docking studies provided by Dr. Guruprasad. The crystal structure used, 2DWB,<sup>117</sup> lacks the full activation loop. It is also not in phosphorylated (activated) form. It also appeared that one of the docking solutions of CoA was used as a template to dock potential lead compounds. This assumption reduced the accuracy of computational docking results significantly. As such, re-examination of the docking results were carried out. The results showed that for the proposed synthetic candidates, **1** and **2**, by Dr. Guruprasad, none of the scored docking poses indicated a good fit of the analogues inside the ATP binding site. For compound **1**, the docking results showed that none of the 10 docking solutions showed a conformation where the heteroaromatic head group sat inside the hinge region. The highest scored docking solution (88.0) was in opposite direction to CoA. (Figure 2.24)



**Figure 2.24.** Binding poses of compound **1** in AURKA ATP binding site suggested by GOLD. Highest scored docking pose (88.0) of **1**. ADP (C atoms in grey, surface area shown). **1** (C atoms in green).

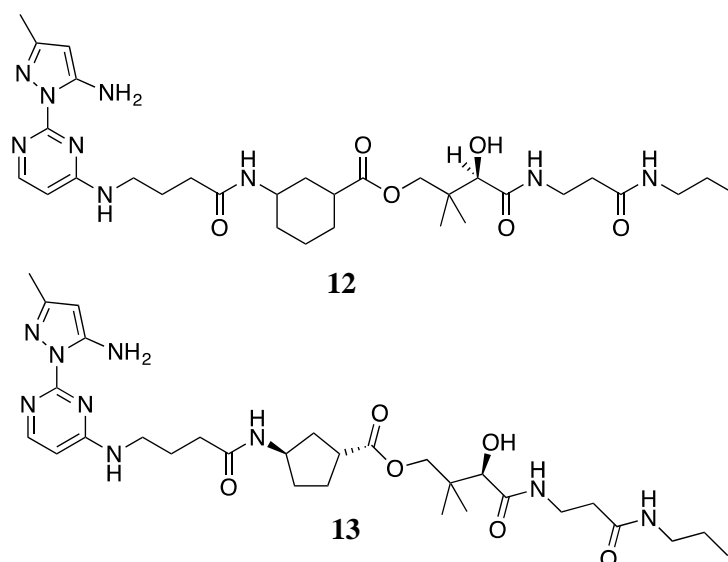
For compound **2**, the docking results showed that the molecule adopted a conformation where the heteroaromatic head group was completely out of the binding hinge region (Figure 2.25). None of the 10 docking solutions showed a conformation where the heteroaromatic head group sat inside the hinge region in similar manner as the core of VX680 (Figure 1.10).





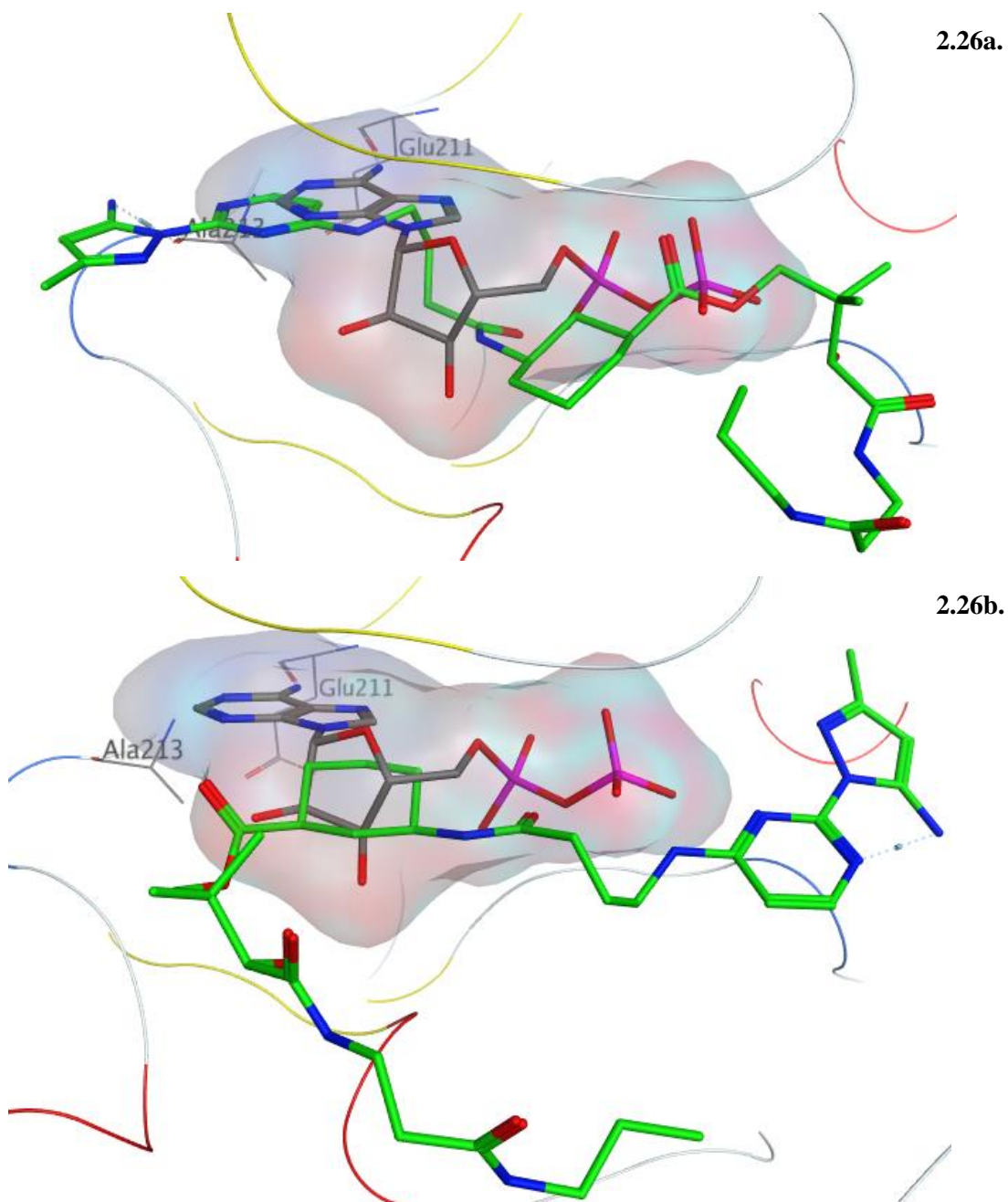
**Figure 2.25.** Binding poses of compound **2** in AURKA ATP binding site suggested by GOLD. Highest scored docking pose (90.2) of **2**. ADP (C atoms in grey, surface area shown). **2** (C atoms in green).

In addition, synthesis of the head group of proposed synthetic candidates **1** and **2** could not be achieved. Instead, the heteroaromatic head groups provided by my colleague allowed the synthesis of **12** and **13**.

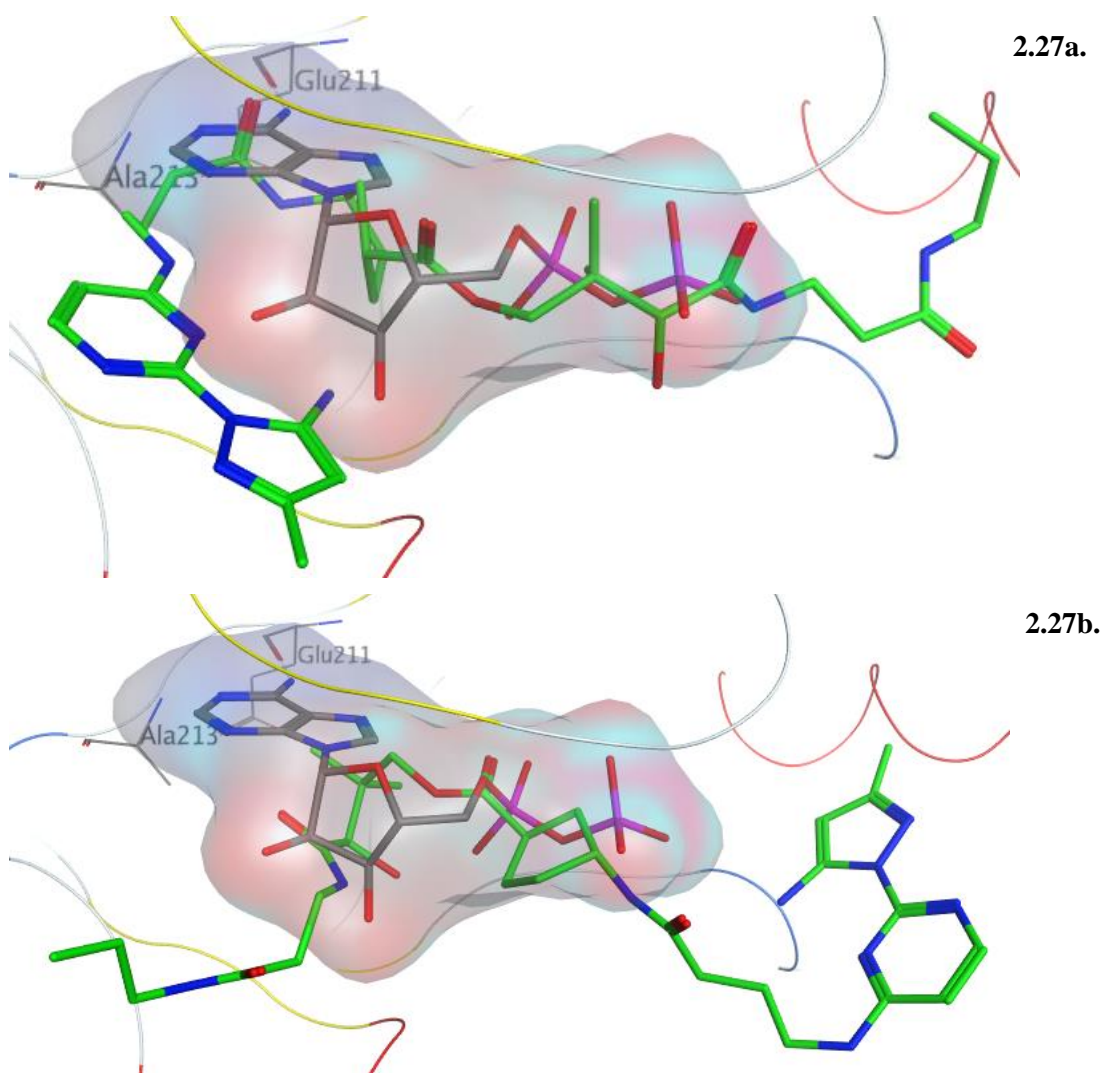


Previous crystal structure of VX680 demonstrated that the N-1 and N-2 of the pyrazole ring were responsible for H-bonding with the hinge region. (Figure 1.10) As a result, it casts doubt on the bindings of compound **12** and **13** in AURKA catalytic domain. Nevertheless the computational dockings were re-examined for possible conformations of **12** and **13** in the binding site. For analogue **12**, the suggested conformations of the ligand either came out of the hinge region or in opposite direction of ADP. (Figure 2.26) Figure 2.26a showed the highest

scored docking pose of **12** where the aromatic head came out of the hinge region. It was the closest distance to the hinge region suggested by all 10 poses. Other poses were either further away from the binding site or in the complete opposite direction as in Figure 2.26b. For analogue **13**, similar poses were observed. The compound was predicted to not fit in the hinge region. Similar to analogue **12** it either came out of the pocket or in opposite direction of ADP. (Figure 2.27) The docking results were in agreement with later preliminary IC<sub>50</sub> results, which of **12** is >7 mM and of **13** is >2 mM. For this reason, the analogues were not pursued further in synthesis.



**Figure 2.26.** Binding poses of CoA's analogue **12** in AURKA ATP binding site suggested by GOLD. a) Highest scored docking pose (85.6) of **12**. b) Third highest scored docking pose (77.7) of **12**. ADP (C atoms in grey, surface area shown). **12** (C atoms in green).



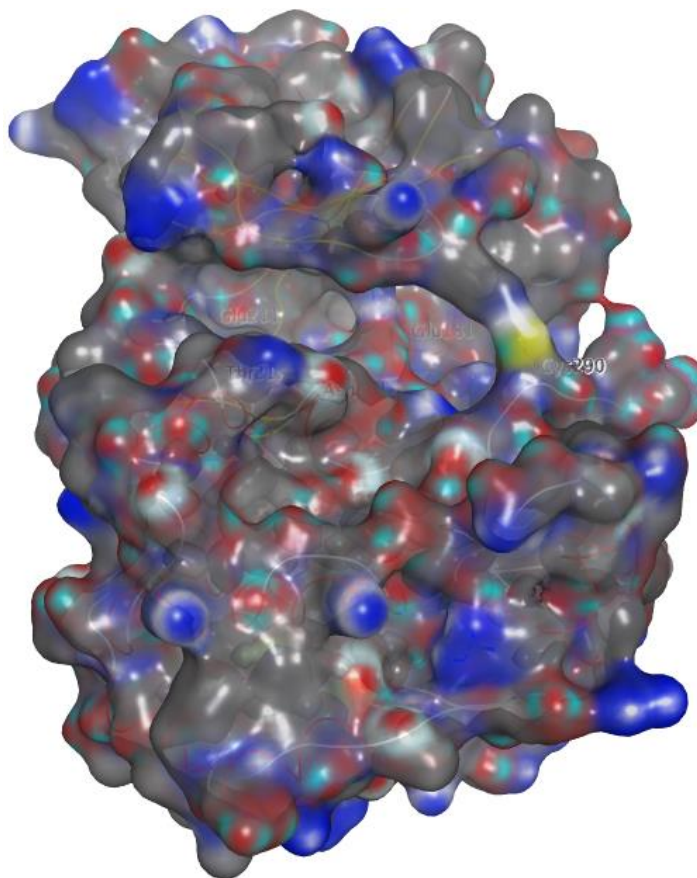
**Figure 2.27.** Binding poses of CoA's analogue **13** in AURKA ATP binding site suggested by GOLD. a) Highest scored docking pose (95.2) of **13**. b) Second highest scored docking pose (91.0) of **13**. ADP (C atoms in grey, surface area shown). **13** (C atoms in green).

#### 2.4.2. Methods of new design for the 2<sup>nd</sup> generation of compounds

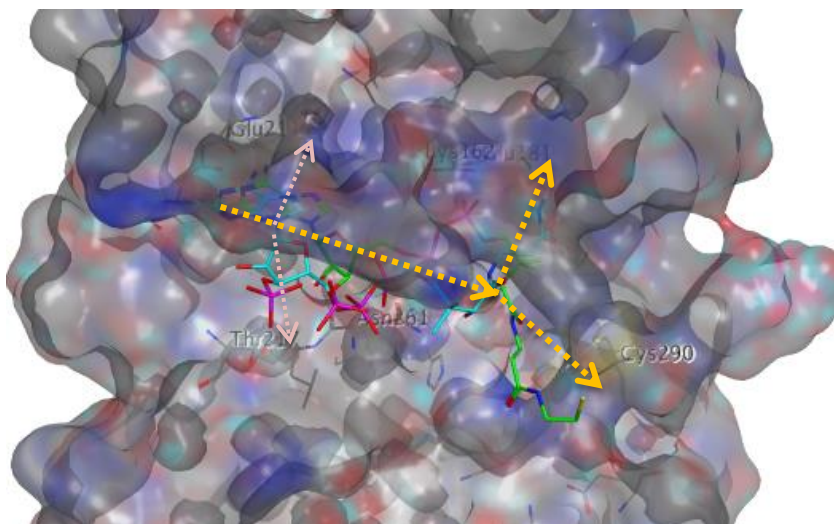
As briefly stated in Chapter 1, the first generation of compounds was halted after the thorough review. In Section 2.4.1, I showed that the compounds recommended for synthesis by Dr. Guruprasad in the first generation of compounds were invalid candidates as AURKA inhibitors. Following the “proof of concept” direction (Section 1.5.2) I outlined the following computational method to design molecules for the purpose of validating biochemical and computational findings of CoA and to understand in detail the catalytic domain of AURKA. These are to set the foundation for designing an AURKA selective inhibitor. The method consists of 4 steps. Firstly, the surface area of AURKA (PDB: 1OL7) (Figure 2.28a) was examined. Crystal structures of known inhibitors were superimposed to examine ligand-receptor interactions (mostly around hinge region). The possible binding modes (with and without the correction of C290 sidechain) of CoA were superimposed to explore ligand-receptor interactions in the part that had not been exploited. (Figure 2.28b) After that, a part of



the ATP binding pocket was chosen to further design structure and investigate SAR. This was followed by screening an in-house database of inhibitor scaffolds of kinases to choose a scaffold to build up new structures that may offer desired ligand-receptor interactions. Finally, the designed structures were docked and evaluated for further modification/development.



**2.28a. Protein surface map of AURKA generated by MOE.** (PDB: 1OL7). Hydrophobic area (grey), polar areas (red and cyan).

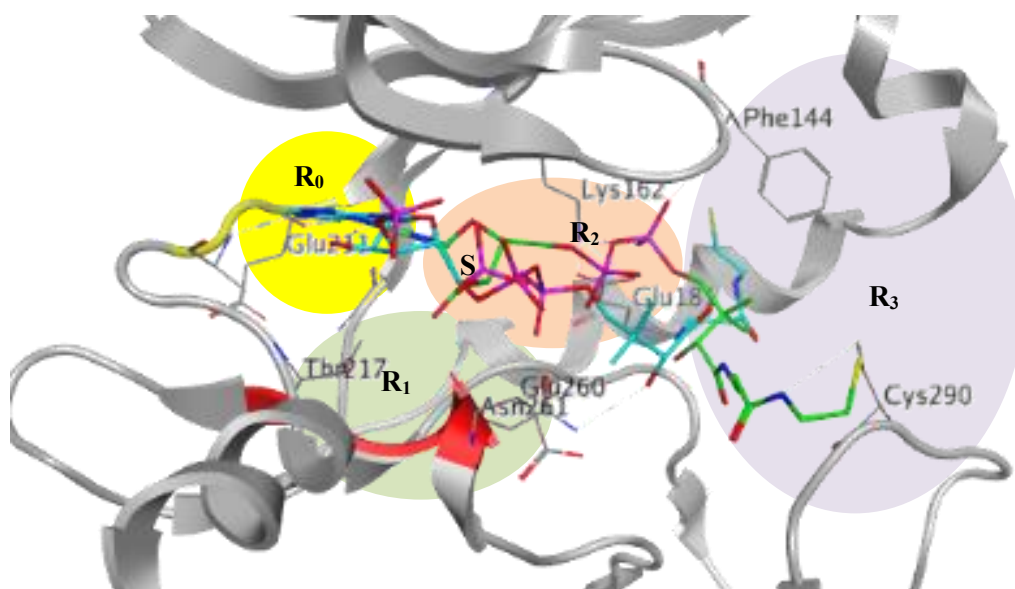


**2.28b. Surface interaction maps of possible binding modes of CoA in AURKA generated by MOE.** CoA binding mode with C290 correction (C $\alpha$  in green). CoA binding mode without C290 correction (C $\alpha$  in cyan). Possible growing directions of new inhibitors: dashed arrows.

**Figure 2.28.** AURKA surface interaction map with and without CoA as ligand.

Examining the surface area of AURKA showed that there are two aspects to consider with regards to the structure of AURKA ATP active site. The first aspect is the hinge region along with small hydrophobic back and front grooves that are commonly known and exploited in the

literature. The second aspect is the extended regions that are mostly unexplored, including: a back groove that was shown to accommodate one binding mode of CoA (C atoms in cyan), demonstrated by an orange arrow in Figure 2.28b; a front, less defined groove containing C290, demonstrated by another orange arrow in Figure 2.28b; and a front polar hump containing T217, E260 and N261. As examined in Figure 1.13, MLN8054 may be able to interact with this region. Taking CoA as the scaffold, the mentioned extended regions of the ATP active site were chosen for further SAR investigation. (Demonstrated by regions R<sub>1</sub>, R<sub>2</sub> and R<sub>3</sub> in Figure 2.29).



**Figure 2.29.** AURKA ATP binding site divided into areas of investigation. R<sub>0</sub>: head group bound to hinge region, S: ribose region; R<sub>1</sub>: polar hump containing T217, E260 and N261; R<sub>2</sub>: phosphate binding site; R<sub>3</sub>: pantetheine binding region. (PDB: 1OL7)

Due to time limitation and for the purpose of comparison, it was decided that R<sub>0</sub> and S were first kept the same as in CoA. In other words, the first new designed compounds (the second generation of compounds) would contain an adenosine head group.

#### **2.4.2.1. Exploring the phosphate binding region, R<sub>2</sub>**

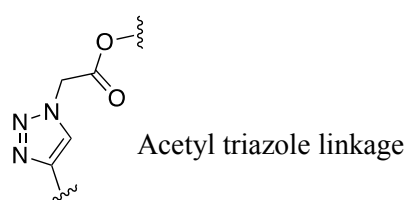
The phosphate binding region, linking the ribose and pantetheine region was first investigated. The idea was to choose a linker that is chemically robust and non-toxic to replace the pyrophosphate. The properties of being chemically robust is important as it allows generation of a large set of compounds in a short period of time, important for SAR study. Literature has suggested that triazole systems can be used as a phosphate mimic.<sup>118</sup> When applying this to a pyrophosphate system, there are two concerns:

- (i) The length and conformational flexibility of this chemical linkage compared to the pyrophosphate system.
- (ii) The reduction in the number of intermolecular hydrogen bonding compared to that provided by the pyrophosphate system.

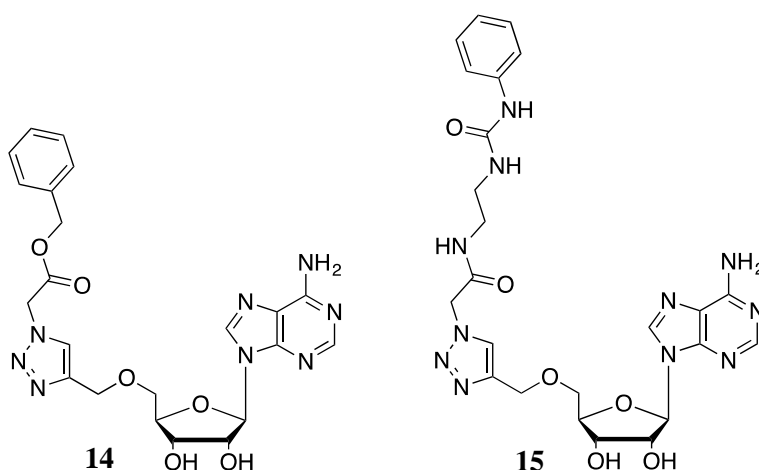


Whether these changes will increase or reduce the activity of CoA/designed compounds against AURKA is open for investigation. However it was decided to start synthesising a set of compounds that structurally differ from CoA in one part but similar in others (e.g. keeping adenosine head and pantetheine tail, replacing pyrophosphate linkage by triazole linkage) for comparison purposes.

In summary, for the second generation of compounds, I chose an acetyl triazole group as the linker replacing the pyrophosphate of CoA. It should be noted that the acetyl triazole group is about 1.5 Å longer than the pyrophosphate. The length difference may not be significant for short molecules but may play a larger role in long molecules where the upper limit for the length to fit in the binding site is reached.

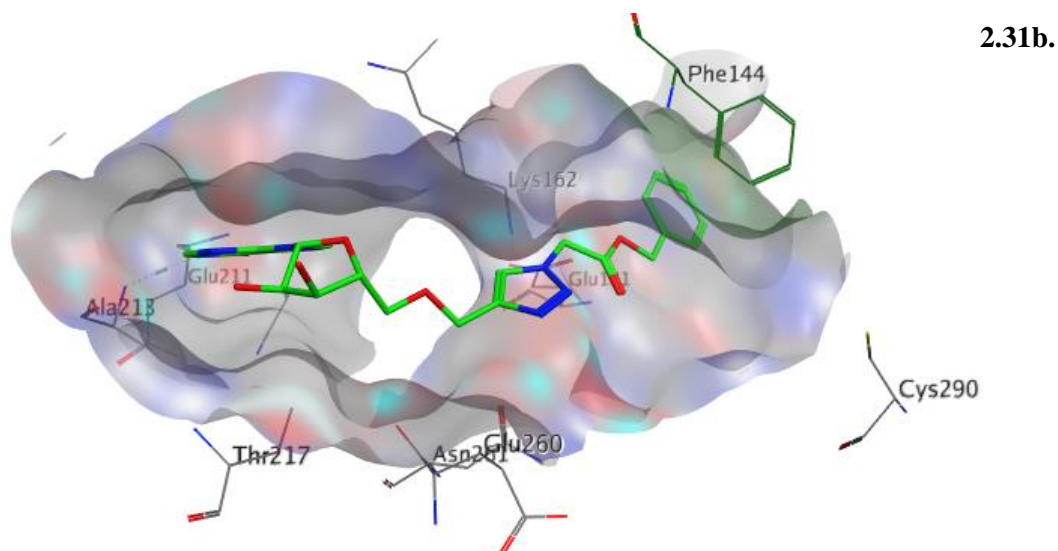
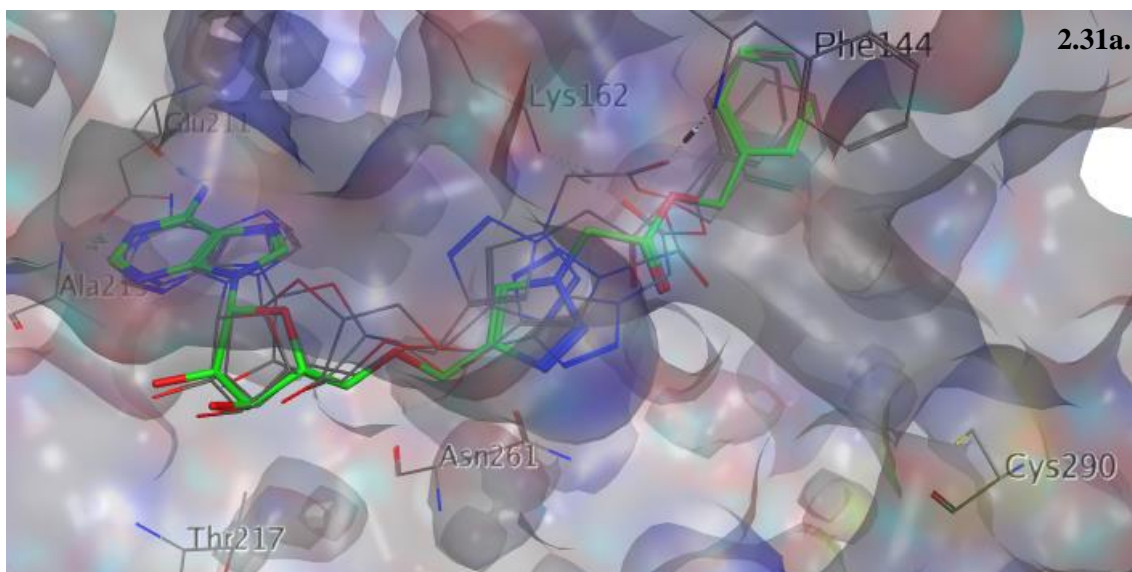


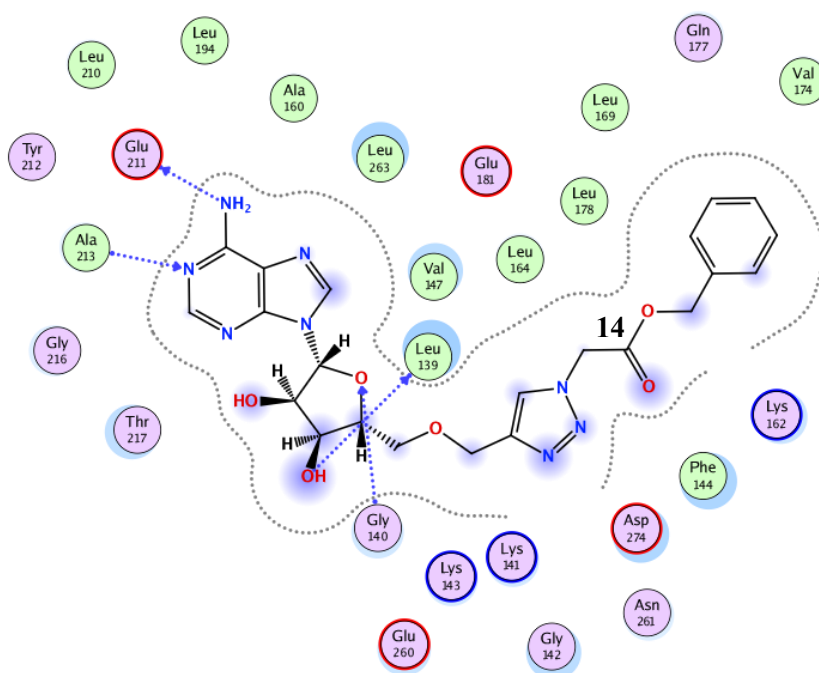
Through further in-house database<sup>119</sup> screening of inhibitor scaffolds of kinases and docking study using GOLD of several structures, it was realised that compounds with a phenyl group at the end with appropriate length could form  $\pi$ - $\pi$  stacking interactions with F144 of AURKA ATP binding site. This interaction was favourable and molecules always seemed to rearrange themselves to accommodate it. F144 is the residue guarding the end of the phosphate binding region and the beginning of the pantetheine binding region. As a result, two compounds **14** and **15** was proposed for docking and synthesis. (Figure 2.30) The synthesis was completed by me.



**Figure 2.30.** Designed compounds to probe the interactions in the phosphate binding region. CHEMPLP suggested that six out of seven solutions showed a consistent binding mode in which the phenyl group was stacked under F144 residue. The acetyl triazole linker might form H-bonds with K162 to further secure the position as suggested by poses. Changing the -OBn end to -OMe resulted in a loss of consistent prediction of its conformation and much lower overall scoring. However **14** was not soluble in water. In order to improve the solubility of **14**,

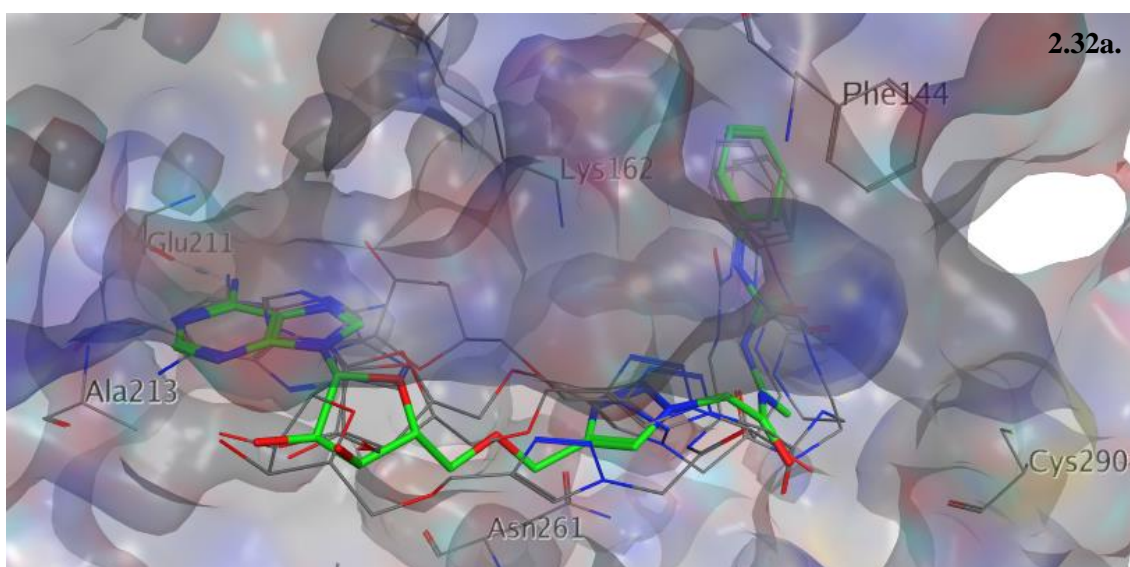
**14** was modified to **15** (containing a urea group), which was later found to be soluble in water as predicted. Besides, all of the more successful Aurora kinase inhibitors present did not contain an ester bond. **15** was designed to contain no ester bond.

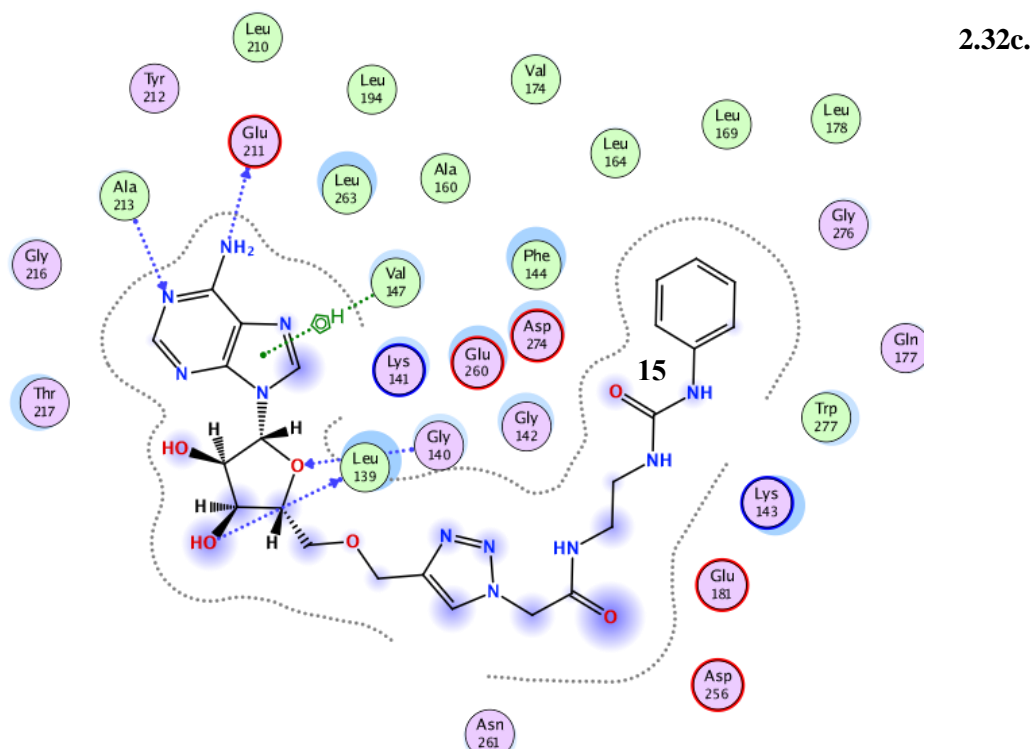
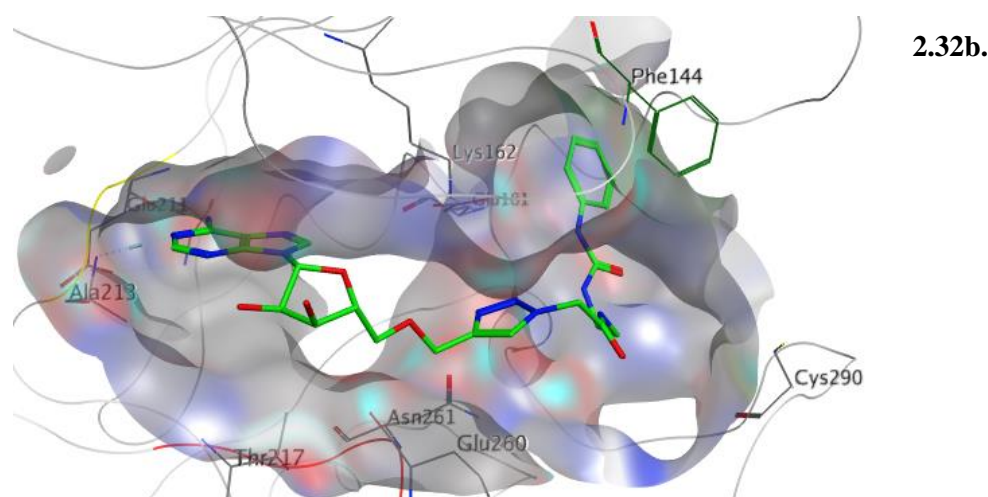




**Figure 2.31.** Conformation of **14** in AURKA ATP binding site. a) Scored docking of **14** in AURKA generated by GOLD. Highest scored docking pose (C atoms in green). Other scored docking poses (C atoms in grey). b) Highest scored docking pose (96.4) of **14**. c) Ligand interaction maps of best scored docking pose of **14** in AURKA generated by MOE.

Docking results of **15** showed comparable scoring to **14**, all the valid solutions (8 out of 10 generated solutions) suggested consistent conformations where the phenyl ring was placed under F144. (Figure 2.32)



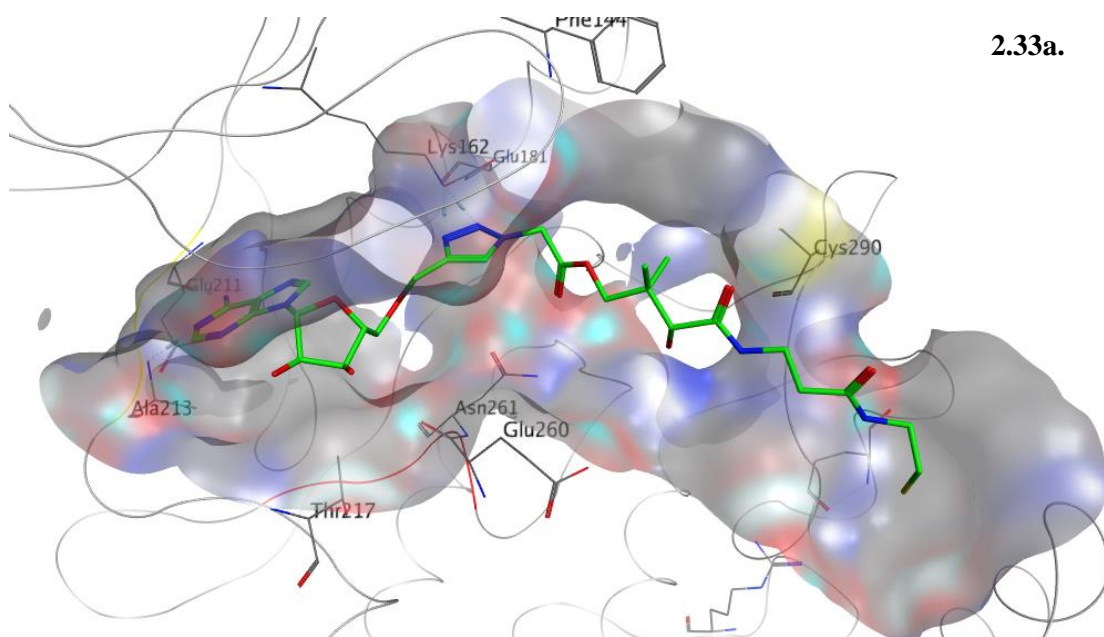


**Figure 2.32.** Conformation of **15** in AURKA ATP binding site. a) Scored docking of **15** in AURKA generated by GOLD. Highest scored docking (C atoms in green). Other scored docking poses (C atoms in grey). b) Highest scored docking pose (99.1) of **15**. c) Ligand interaction maps of best scored docking pose of **15** in AURKA generated by MOE.

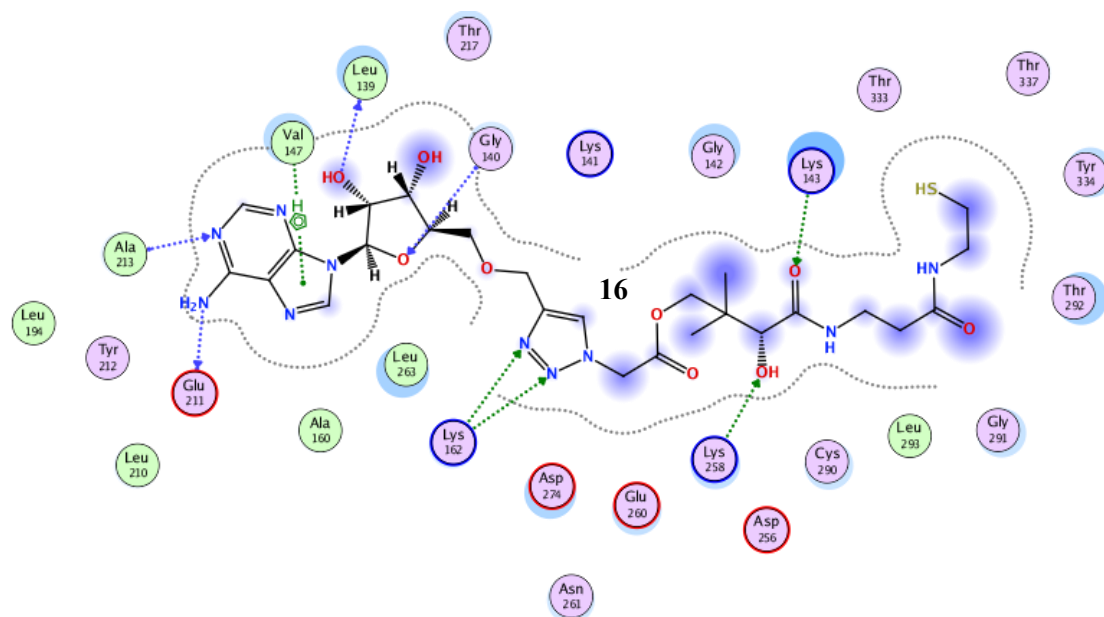
#### 2.4.2.2. Exploring the pantetheine binding region, $R_3$

Using adenosine as the head group, and acetyl triazole as the linker to replace the pyrophosphate group, it would be informative to construct a pantetheine analogue group to compare the possible effect of the pantetheine group on the overall inhibitory activity of CoA and to explore the pantetheine binding site for further drug development. Compound **16** was designed based on this idea. (Figure 2.33) Docking results showed only 30% - 40% chance of **16** to be fit in the hinge region and the -SH was not in as close proximity with C290 as that of CoA. The result suggested there were some good interactions between the triazole and K162 so

it was suggested that the length of the tail should be cut down by 2 CH<sub>2</sub> groups or the –CH<sub>2</sub>COO group linked between the triazole and the tail needed to be removed.



2.33a.

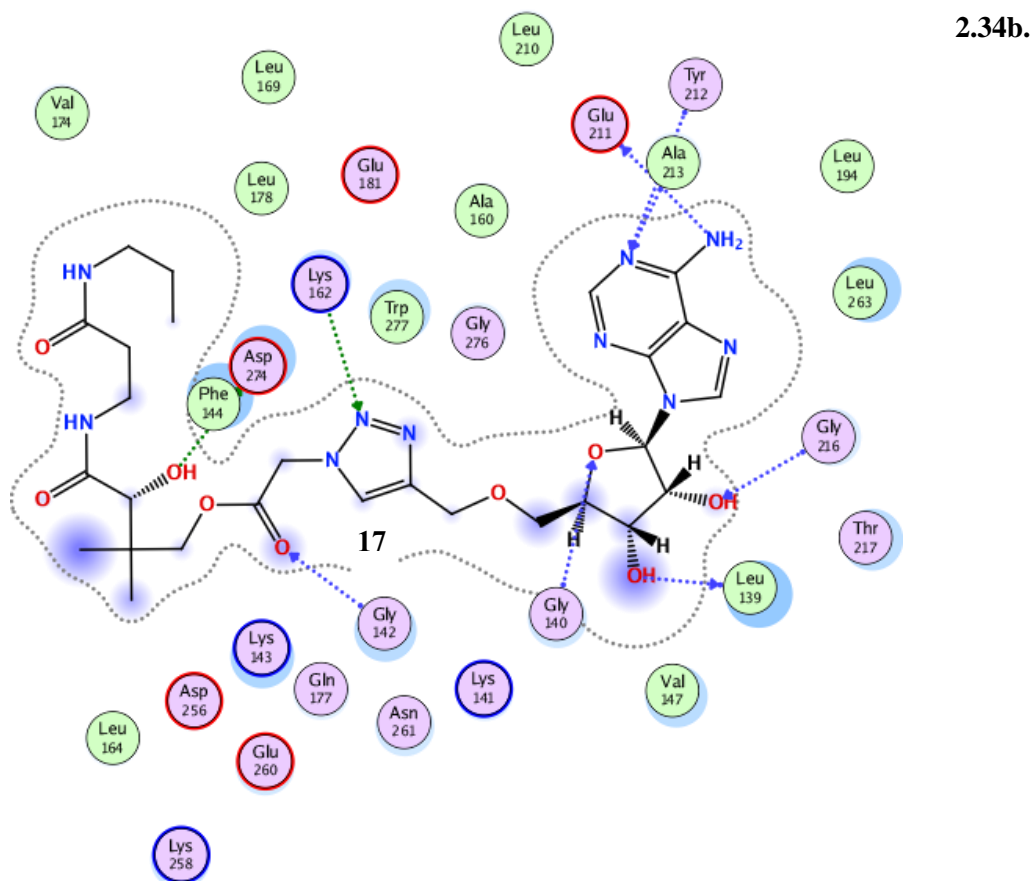
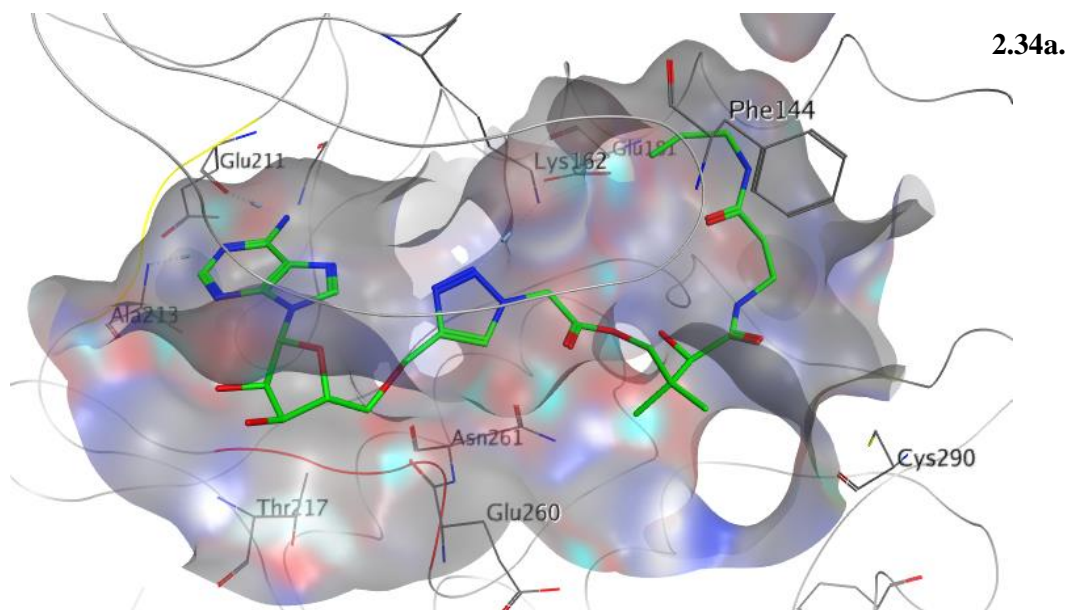


2.33b.

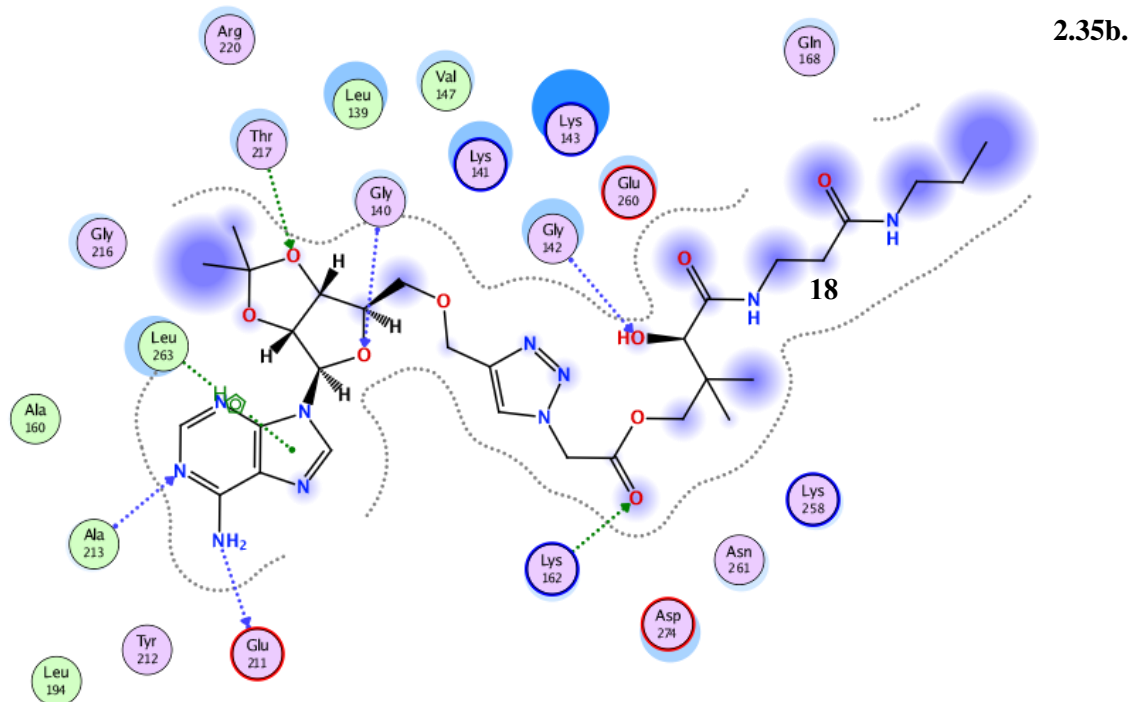
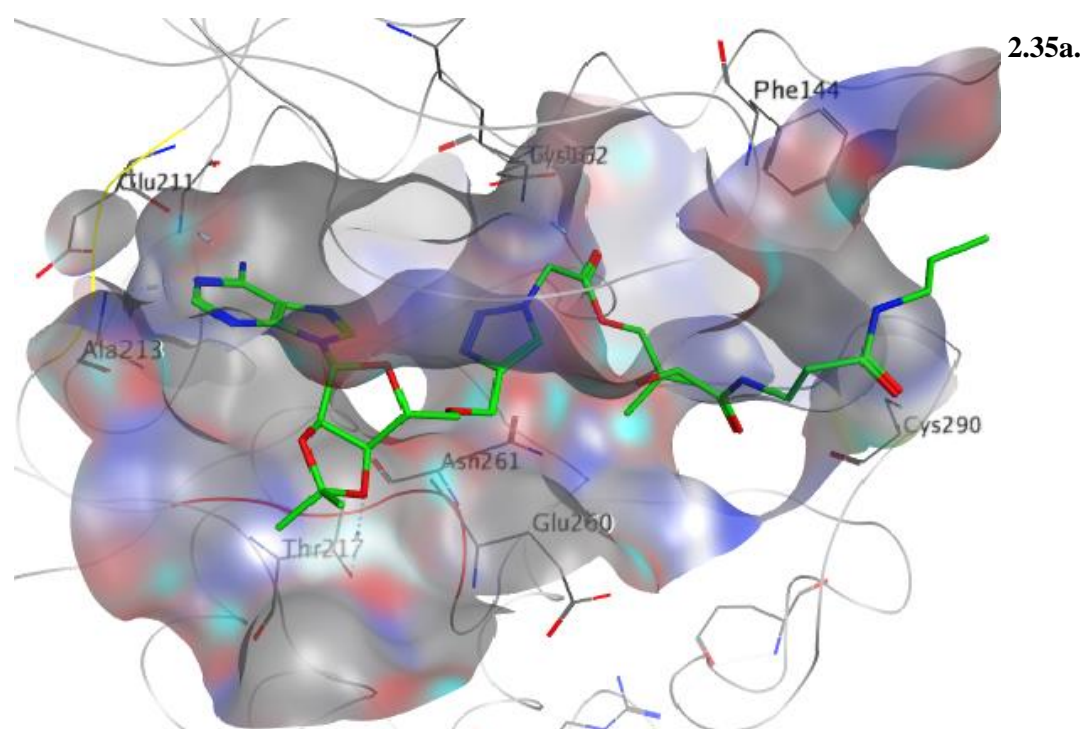
**Figure 2.33.** Conformation of **16** in AURKA ATP binding site. a) Best scored docking pose (91.7) of **16** (C atoms in green). b) Ligand interaction maps of best scored docking pose of **16** in AURKA generated by MOE.

The effect of –SH group was investigated by designing **17**, which is similar to **16** but the –SH is replaced by –CH<sub>3</sub> group. **18** was proposed to be synthesised to investigate the effect of polar groups at the ribose. This will help deciding how to explore R<sub>1</sub> region later. Docking suggested that there was only 30% of finding **17** inside AURKA ATP binding site. (Figure 2.34) The docking result for **18** showed that if the acetyl was introduced to protect –OH groups of the ribose, there would be little chance of finding **18** inside the binding site (<10%) hence no inhibition. (Figure 2.35)





**Figure 2.34.** Conformation of **17** in AURKA ATP binding site generated by GOLD. a) Highest scored docking pose (90.7) of **17** (C atoms in green). b) Ligand interaction maps of best scored docking pose of **17** in AURKA generated by MOE.



**Figure 2.35.** Conformation of **18** in AURKA ATP binding site. a) The single scored docking pose (73.7) suggested **18** in the ATP binding site, not the 5 highest scored docking poses (C atoms in green). Other scored docking poses (C atoms in grey). b) Ligand interaction maps of best scored docking pose of **18** in AURKA generated by MOE.

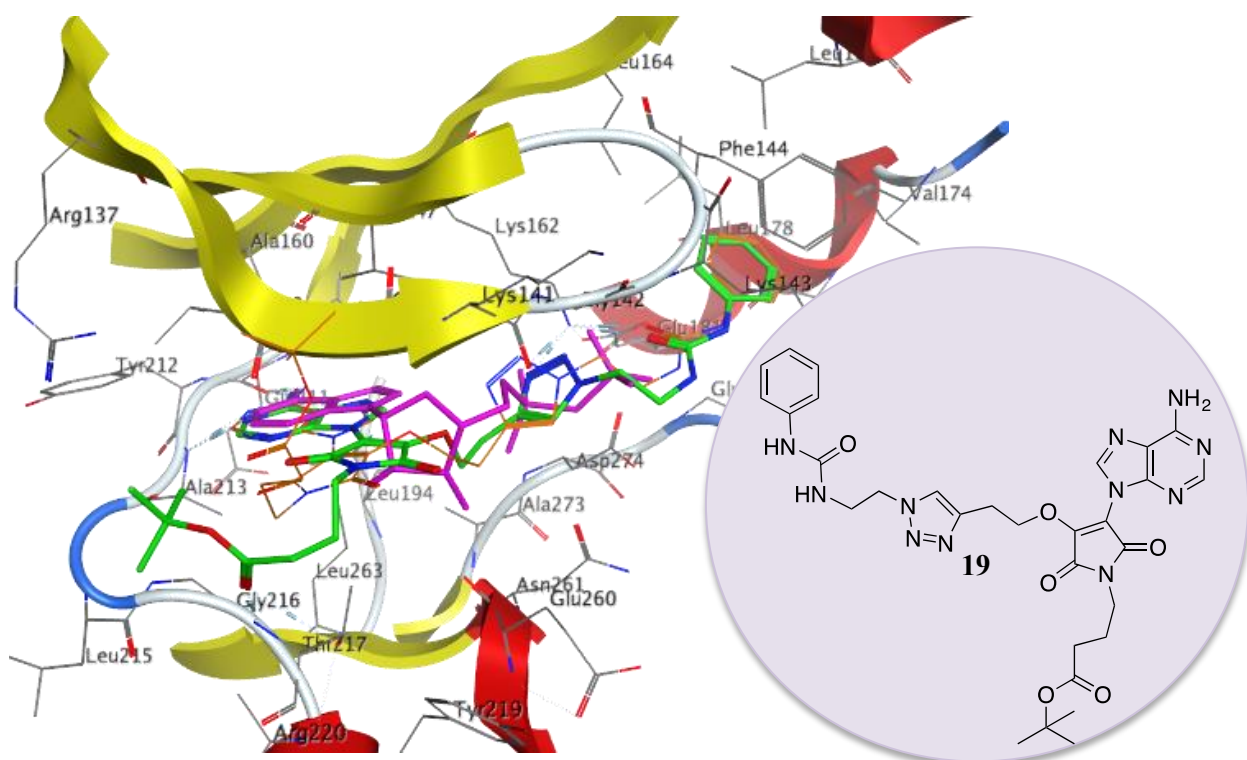
In conclusion, it was suggested from docking results that **14** and **15** would likely show some inhibitory effect in an ATP competitive manner. **16** and **17** may provide some weak inhibition and based on the docking score, **16** may perform a little bit better than **17** but **18** will not exhibit any inhibitory effect. Later, a comparison between the  $IC_{50}$  of these compounds and the docking

result, presented in Section 4.3 would provide a better direction for how to design further compounds for AURKA selective inhibition.

#### 2.4.2.3. Exploring the polar region containing T217, R<sub>1</sub>

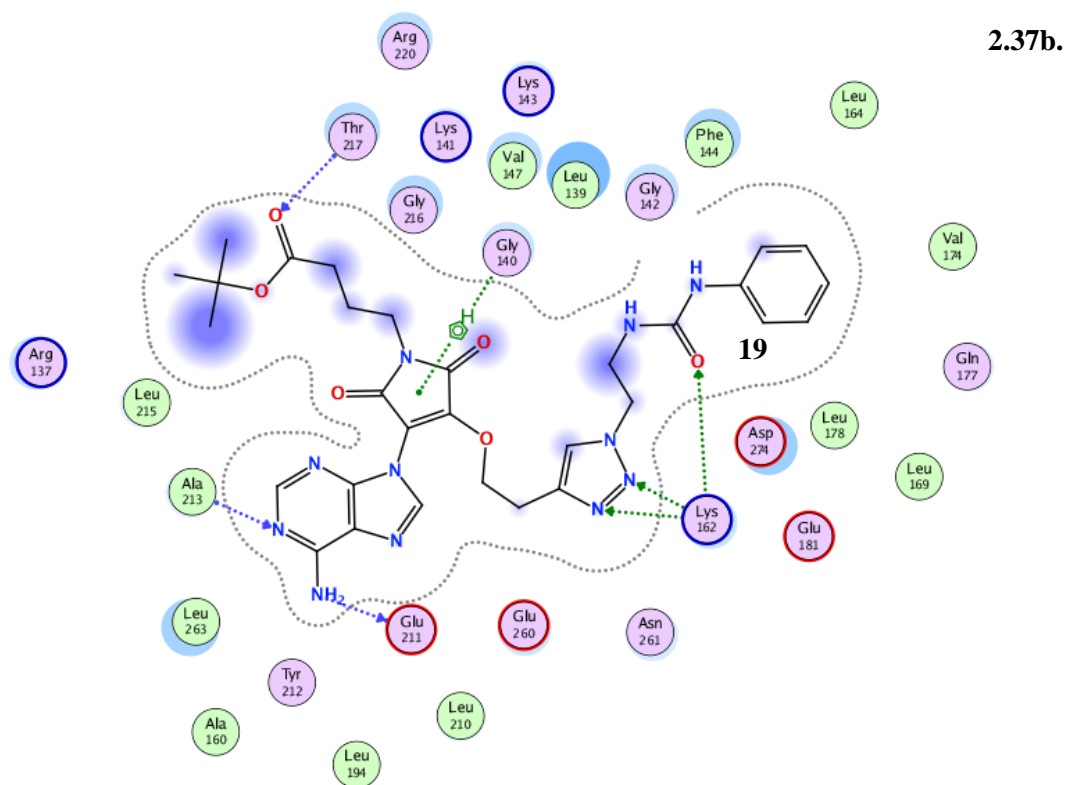
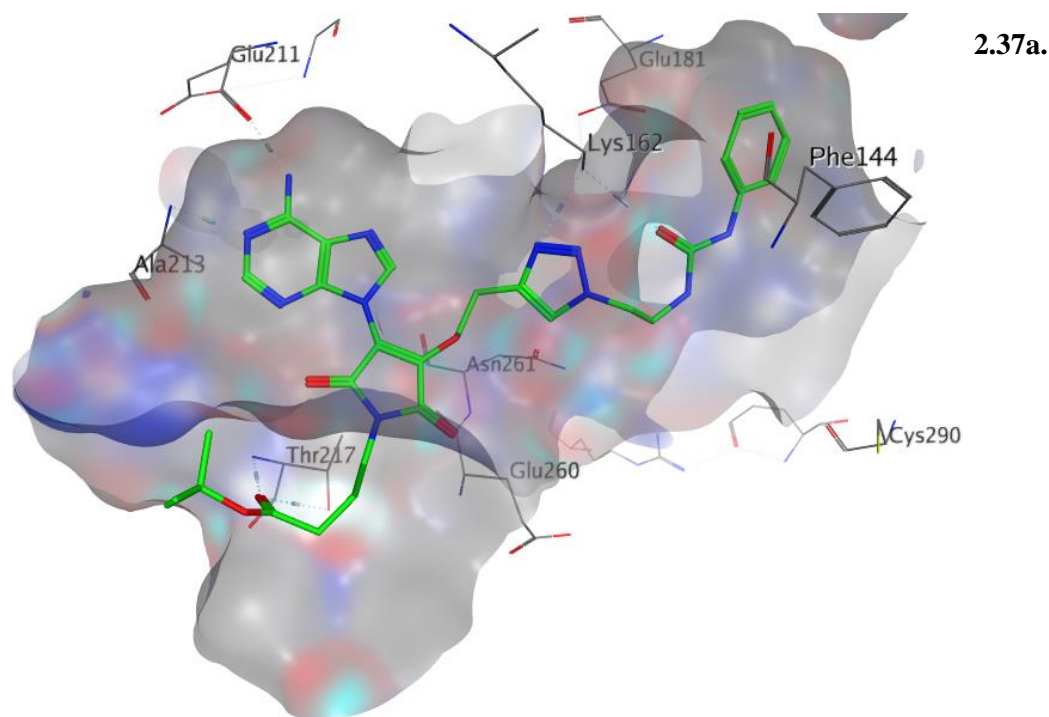
In order to explore the R<sub>1</sub> region, the ribose was replaced by a maleimide moiety. This can act as a linker to link 3 groups: the head group (R<sub>0</sub>), the phosphate mimic (R<sub>2</sub>) and R<sub>1</sub> group. Its chemistry and application has been extensively studied, facilitating the progress of developing the chemistry and generating large library.

Screening a series of structures suggested **19** for initial synthesis, where the adenine head group and the phenyl urea (as in **15**) were kept for the purpose of comparison. Docking results suggest a good inhibition with 70% chance of finding **19** inside the ATP binding site and possible interaction with the polar hump, which contains T217. (Figures 2.36 and 2.37)



**Figure 2.36.** Docking result of **19** in AURKA ATP binding site. Highest scored docking pose (110.1) of **19** in AURKA generated by GOLD. **19** (C atoms in green). ADP (C atoms in pink).

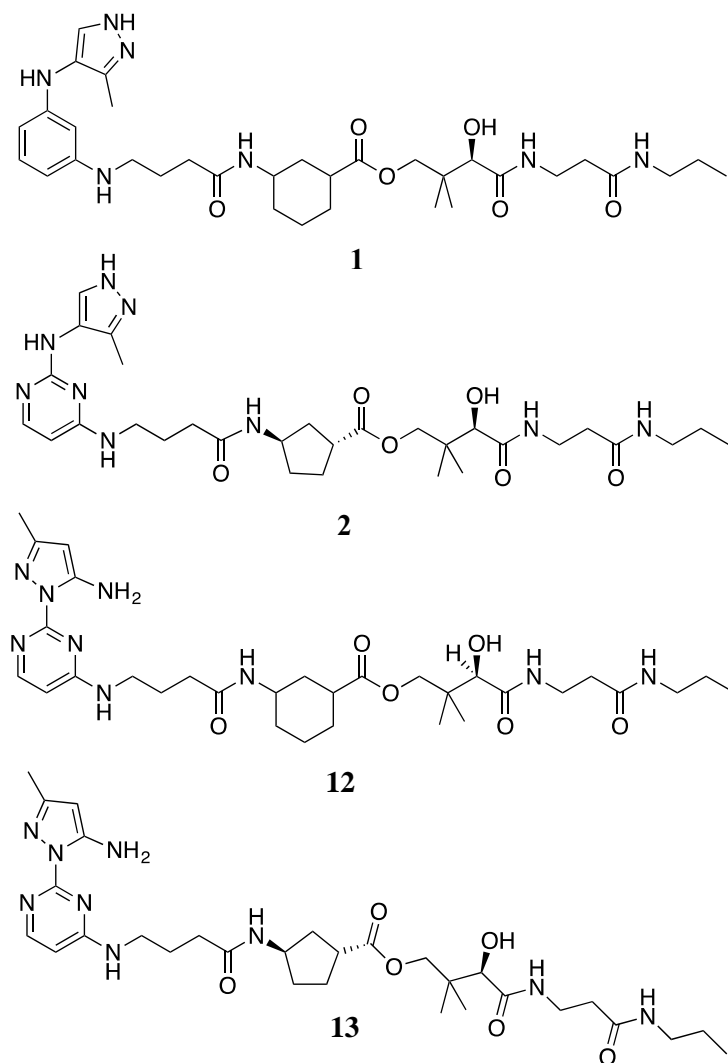




**Figure 2.37.** Conformation of **19** in AURKA ATP binding site generated by GOLD. a) Highest scored docking pose (110.1) of **19** (C atoms in green). b) Ligand interaction maps of highest scored docking pose of **19** in AURKA generated by MOE.

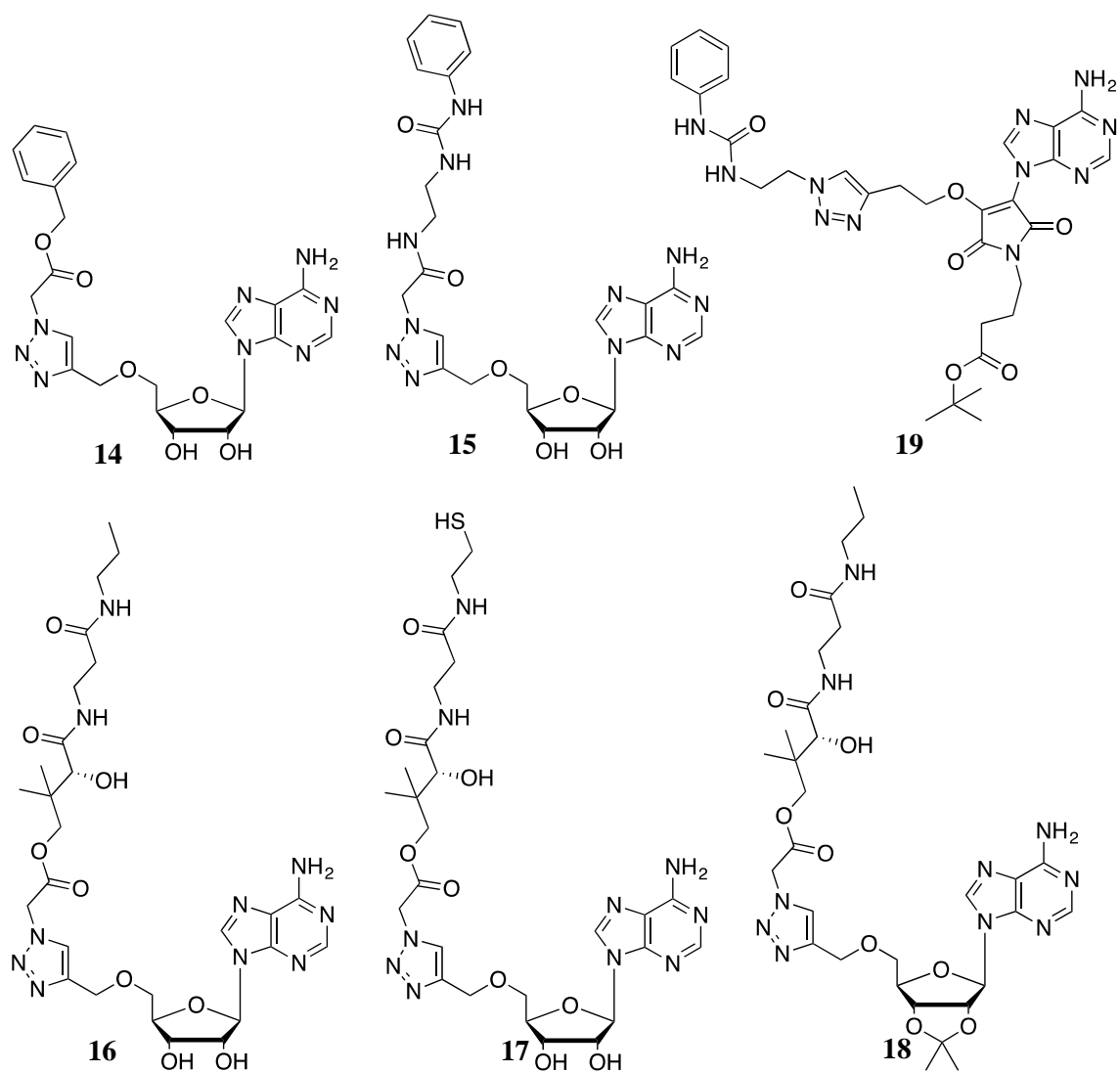
## 2.5. Conclusion

In summary, in approach **A**, which focused on the design and synthesis of an AURKA selective inhibitor based on the structures of VX680 and CoA, (Section 1.5.1) computational work done by Dr. Lalitha Guruprasad suggested four compounds, compound **1**, **2**, **12** and **13** for synthesis. (Figure 2.38) Compound **1** and **2** were unable to be synthesised. The syntheses of **12** and **13** are described in Chapter 3.



**Figure 2.38.** Four main compounds in Approach A.

Further computational work by me showed that none of the four compounds would exhibit any inhibitory effect against AURKA. This resulted in the shift in focus to Approach **B**. The main objective of Approach **B** was to validate the biochemical and computational findings of CoA and to understand in details the catalytic domain of AURKA. These are to set the foundation for designing an AURKA selective inhibitor. Six key designed molecules were synthesised. (Figure 2.39)

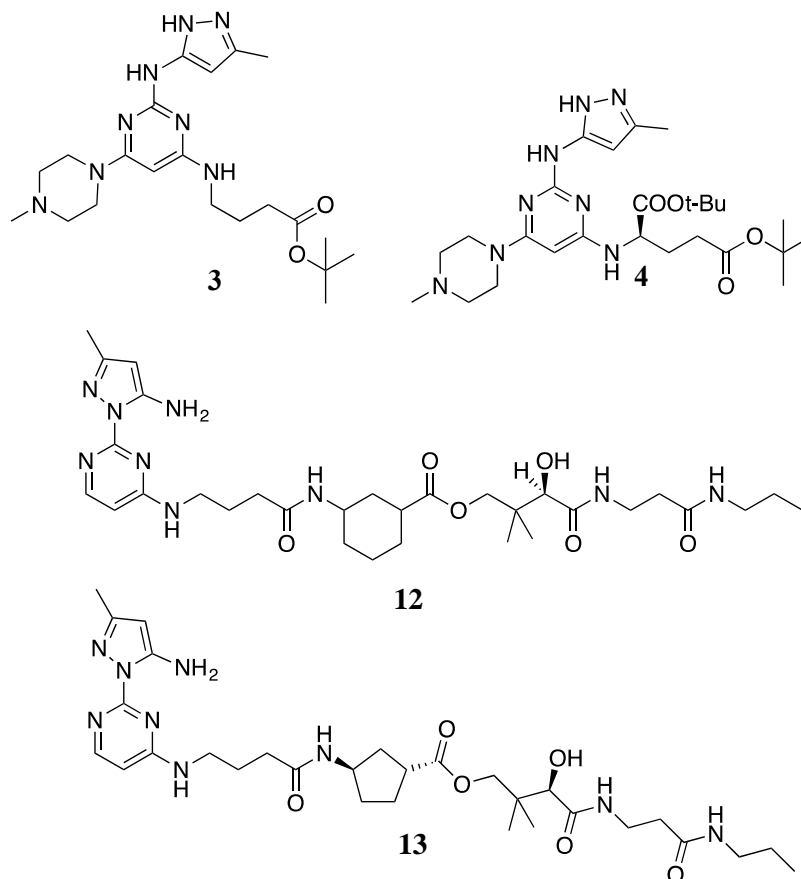


**Figure 2.39.** Six key designed molecules in Approach B – the second generation of compounds.

### 3. Approach A syntheses

#### 3.1. Synthesis of the first generation of compounds

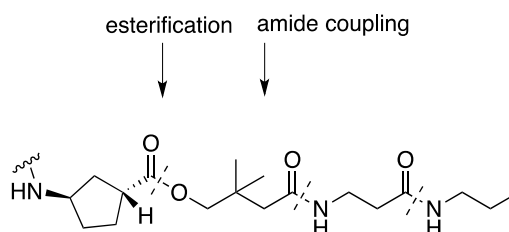
As mentioned in Section 1.5.1 and 2.4, in Approach A, the designed compounds were based on “CoA scaffold” and VX680. This chapter includes the syntheses of compounds **12** and **13**, where the synthetic focus was on the cyclic linker and pantothenamide end, and the initial synthetic attempts of VX680-heteroaromatic analogues, **3** and **4**. (Figure 3.1)



**Figure 3.1.** Synthetic targets of Approach A

#### 3.1.1. Literature research and synthetic plan of the 1<sup>st</sup> generation of compounds

Given an example of the pantetheine-type structure with a linker, one approach to this type of compounds includes amide couplings and esterification reactions. (Figure 3.2)

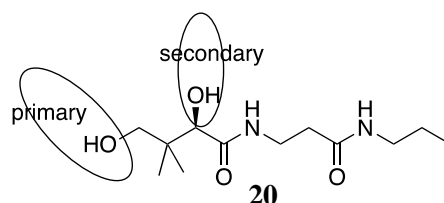


**Figure 3.2.** Synthesis approach to the first generation of compounds

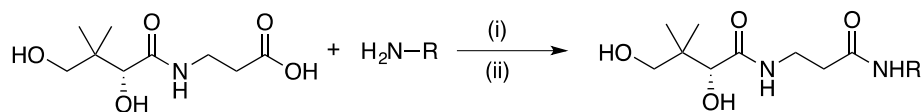
The structure suggests that there are three matters to consider in the synthesis of this part of the designed molecule:

- (i) Amide coupling reaction - Synthesis of pantothenamide tail (i.e. pantetheine analogue)
- (ii) Esterification
- (iii) Possible protecting group problem

*Amide coupling reaction - Synthesis of pantetheine analogues*

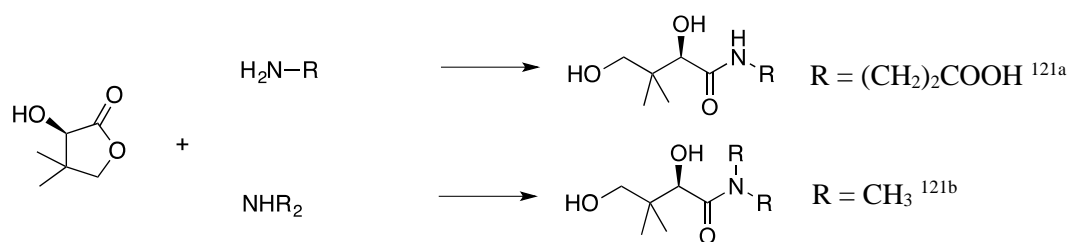


There are two ways to prepare the pantothenamide-type end **20**, either by an amide coupling between pantothenic acid and an amine (Route 1) <sup>120</sup> or by the ring opening of D-pantolactone (Route 2) <sup>121</sup>. Virga *et al.* in 2006 <sup>120</sup> reported the successful syntheses of several pantothenamide analogues by amide coupling between free-acid form of pantothenate and primary amines. (Figure 3.3)



**Figure 3.3.** Amide coupling of pantothenic acid. (i) DPPA, DMF, 0°C; (ii) TEA, 24 h. <sup>120</sup>

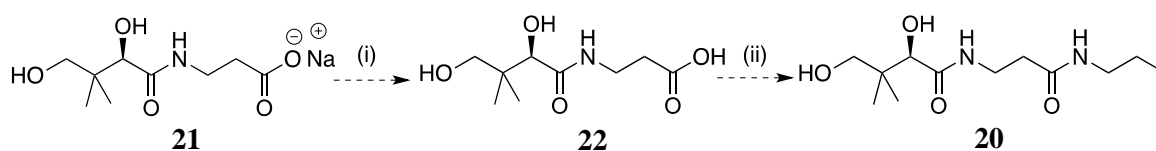
Ring opening of D-pantolactone by amines was reported to be achievable by Kopelevich *et al.* in 1971 <sup>121a</sup> and Hoepfer and Montforts in 1993 <sup>121b</sup>. This was adapted to provide another feasible way of preparing the pantothenamide-type analogues. (Figure 3.4)



**Figure 3.4.** Examples of opening of D-pantolactone by amines. <sup>121</sup>

*Route 1*

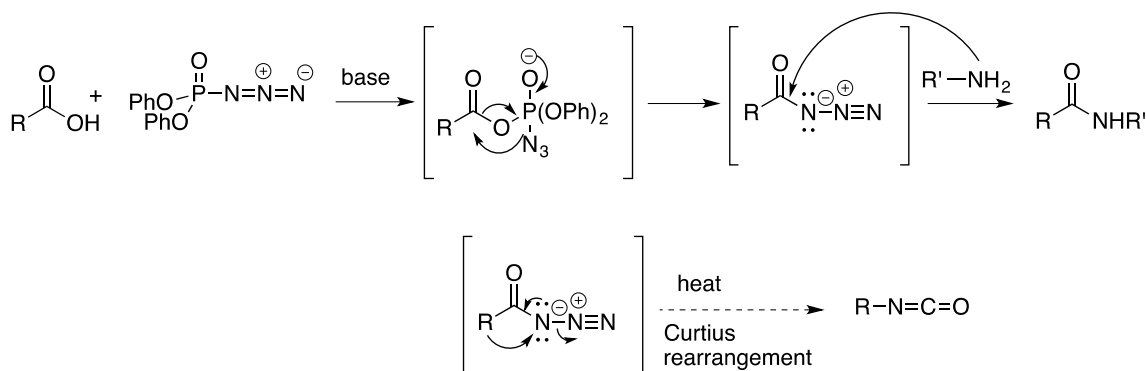
Route 1 (Figure 3.5) adapted from Virga *et al.* <sup>120</sup> started with the free-acid form pantothenic acid **22**. This would provide a quick synthesis to a range of different pantothenamide-type compounds as the free acid can be obtained quickly from the commercially available sodium salts through Amberlite IR-120 resin column.



**Figure 3.5.** Route 1 plan of **20**. (i) Preparation of carboxylic acid in free form; (ii) amide coupling.

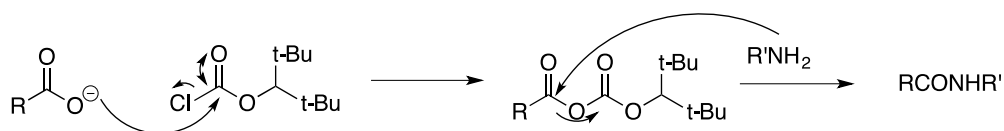
In the reported synthesis by Virga *et al.*, the hydroxyl groups of pantothenic acid **22** were not protected. Due to reactivity and steric reasons, it is likely that significant reaction in esterification at the desired primary alcohol would be achievable without protecting strategy. However, protection at  $-NH_2$  end of the linker is preferred and will be discussed in the next section, linker synthesis. At this point, there were three aspects to be investigated in the effort to make the coupling work: choice of coupling additives, solubility of reagents and/or choice of solvent and temperature.

The amide bond formation can generally be achieved by using coupling additives, generating the active acylating reagent *in situ* in the presence of the amine.<sup>122</sup> A few common coupling reagents were selected. The first one to be considered was diphenylphosphorylazide, DPPA, reported in Virga *et al.* to successfully couple the pantothenic acid to a series of amines. Theoretically this allows the coupling to be achieved in high yield as the acyl azide is relatively unreactive to non-amine nucleophiles and Curtius rearrangement is relatively slow compared to the coupling. (Figure 3.6)<sup>122, 123</sup>



**Figure 3.6.** Amide coupling using DPPA

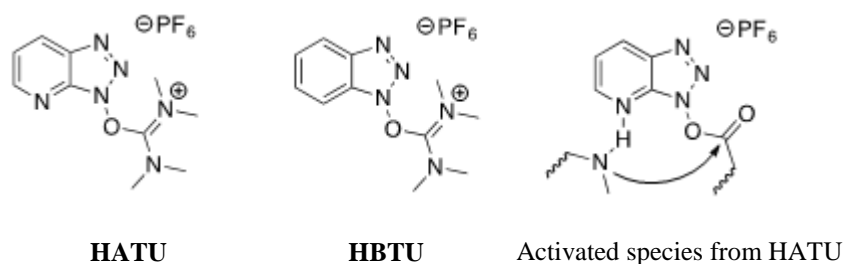
Another selected choice was iso-*tert*-butyl chloroformate, to form an activating anhydride with the starting carboxylic acid. (Figure 3.7) The chloroformate was chosen for reactions where starting materials were soluble in organic solvent.<sup>124</sup>



**Figure 3.7.** Amide coupling using chloroformate

Other choices of consideration were aminium salts, HBTU<sup>125</sup> and HATU<sup>126</sup>, which are commonly used in reactions, where starting reagents are partially soluble in common organic

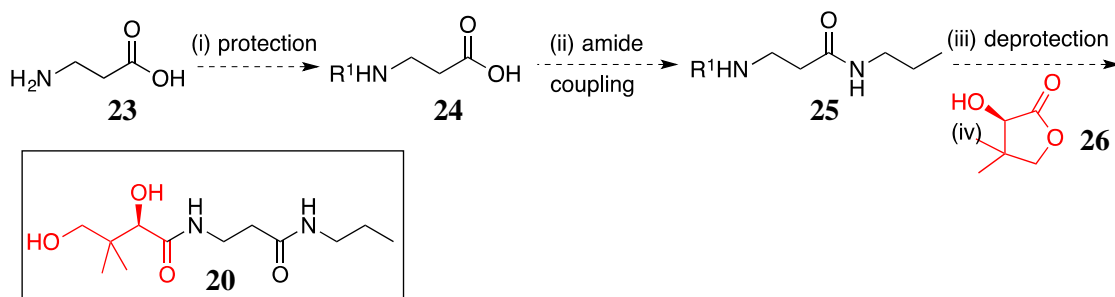
solvents other than DMF. (Figure 3.8) Esterification was first tried out using standard dicyclohexylcarbodiimide, DCC, as the coupling reagent.<sup>127</sup>



**Figure 3.8.** Amide coupling using aminium salts.

#### Route 2

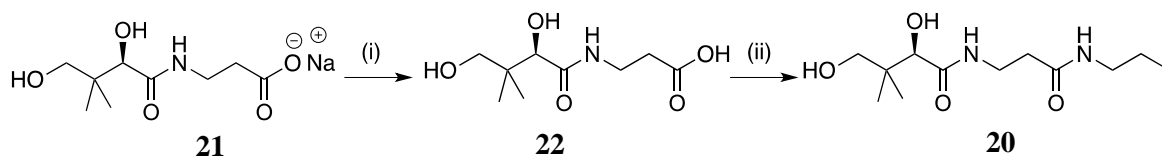
Route 2 uses  $\beta$ -alanine as the starting material followed by an opening of the D-pantolactone ring forming the desired pantothenamide, which then undergoes an esterification, using a coupling agent such as DCC. (Figure 3.9) This approach, if successful, will allow the control of the length between two amide bonds.



**Figure 3.9.** Planned route 2 of **20**. (i) Protection of amino group, R<sup>1</sup>= protecting group; (ii) Amide coupling; (iii) Deprotection; (iv) Ring opening of D-pantolactone.

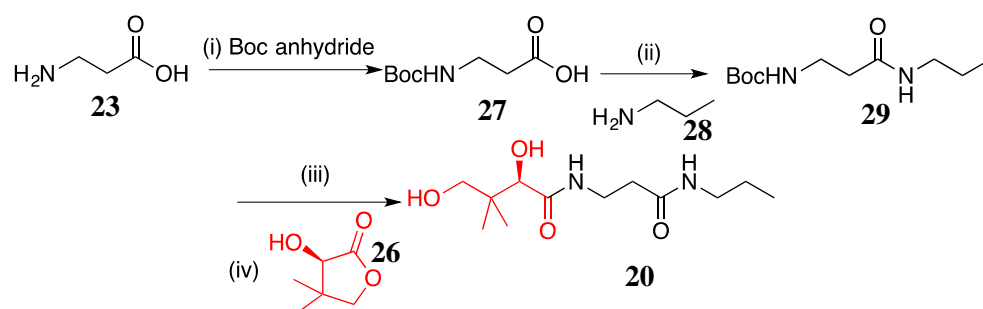
#### 3.1.2. Synthetic results of pantothenamide tail

In synthetic route 1, which was adapted from Virga *et al.* (Figure 3.10), the free pantothenic acid can be obtained quantitatively by passing the commercially available sodium salt through Amberlite IR-120 resin column. The free acid form **22** was then coupled to propylamine using DPPA in the presence of triethylamine in dry DMF. Compound **20** was obtained successfully in 96% yield in 2 g scale.



**Figure 3.10.** Synthetic route 1 of **20**. (i) Amberlite IR-120 resin, MeOH, water, quantitative; (ii) DPPA (1.5 eq), TEA (2.1 eq), DMF, 0°C, 45 min, propylamine (1.6 eq), 0°C to r.t., 20 h, 96%.

In synthetic route 2 (Figure 3.11), the starting  $\beta$ -alanine was Boc-protected using Boc anhydride in sodium hydroxide and water with modified conditions from the literature to give *N*-Boc  $\beta$ -alanine in 57% yield.<sup>128</sup> Amide coupling of *N*-Boc  $\beta$ -alanine to propylamine was achieved using isobutyl chloroformate to yield **29** in 58% yield as a white powder.<sup>129</sup> Deprotection of the Boc group by TFA in  $\text{CH}_2\text{Cl}_2$  provided the deprotected form of **29** quantitatively as a colourless oil. The product was then used directly to open the D-pantolactone ring using a procedure adapted from Kopelevich *et al.*<sup>121a</sup> to obtain **20** as a colourless oil (total yield of 30% through 4 steps).



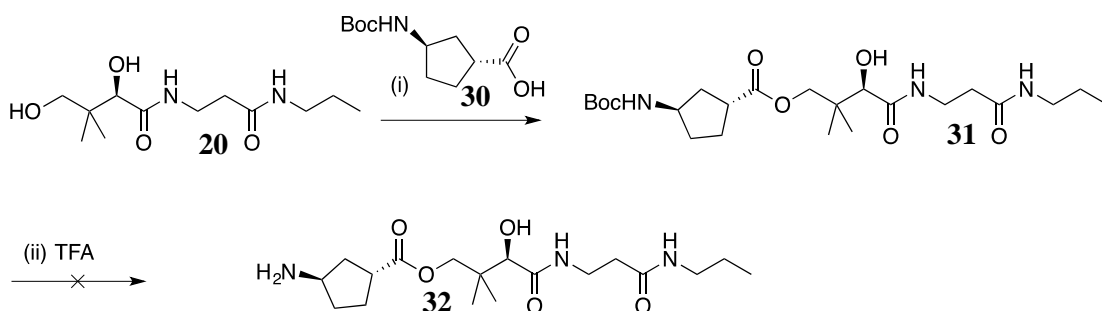
**Figure 3.11.** Synthetic route 2 of **20**. (i) Boc anhydride (1.0 eq), NaOH, water, r.t, 22 h, 57%; (ii) iso-*t*-butyl chloroformate (1.2 eq), **28** (1.3 eq),  $\text{NEt}_3$  (1.4 eq), THF, 0°C to r.t, 20 h, 58%; (iii) TFA:  $\text{CH}_2\text{Cl}_2$  = 1:1, 15°C, 6 h, quantitative; (iv) **26** (1.0 eq),  $\text{NHET}_2$  (xs), MeOH, reflux, 29 h, 92%.

### 3.1.3. Synthesis of linkers and esterification

#### 5- and 6-membered ring linkers

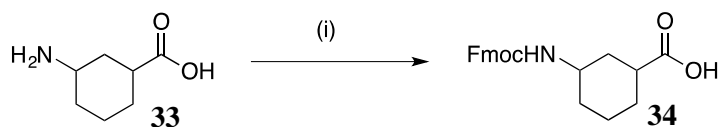
The first generation of compounds, **12** and **13**, consists of a 5- or 6-membered ring linker. (Figure 3.1) Firstly an esterification was performed between (*R*)-2,4-dihydroxy-3,3-dimethyl-*N*-(3-oxo-3(propylamino)propyl)butanamide, **20** and commercially available (*1R, 3R*)-*N*-BOC 1-aminocyclopentane 3-carboxylic acid, **30** using DCC, DMAP in THF, 45°C for 15 h. It was successful, yielding **31** (33% yield). However, later attempts of removal of the -Boc group by TFA and acidic resin Amberlyst-15 were unsuccessful. Rubin in 1948 reported that sodium pantothenate stability dropped 40% after 48 h at 45°C in pH 4.<sup>130</sup> As such it was suspected that the pantothenamide tail might not be able to withstand strong acidic or basic conditions or high temperature. Consequently, it was envisaged that the Fmoc group, which can be removed under mild basic condition, would be more suitable. The temperature recorded (68°C, reflux condition) for esterification was reexamined and shown to be incorrect due to heating plate error. Several attempts showed that at temperature higher than 50°C, the pantothenamide, **20**, degraded overnight. The esterification was successfully reproduced at around 45°C.





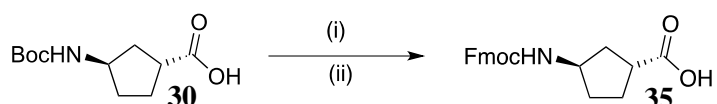
**Figure 3.12.** Attempted synthesis of **32**. (i) DCC (2.0 eq), DMAP (0.4 eq), THF, 0°C to r.t, 2 h then **30** (1.0 eq), 45°C, 15 h, 33%; (ii) TFA, DMF.

3-(((9*H*-Fluoren-9-yl)methoxy)carbonyl)amino)cyclohexane-1-carboxylic acid, **34** was synthesised from the commercially available 3-aminocyclohexane-1-carboxylic acid (*cis*- and *trans*- mixture with >95% *cis*-3-aminocyclohexane-1-carboxylic acid). The 6-membered ring linker was initially used as a model for the 5-membered ring linker due to competitive pricing and availability.



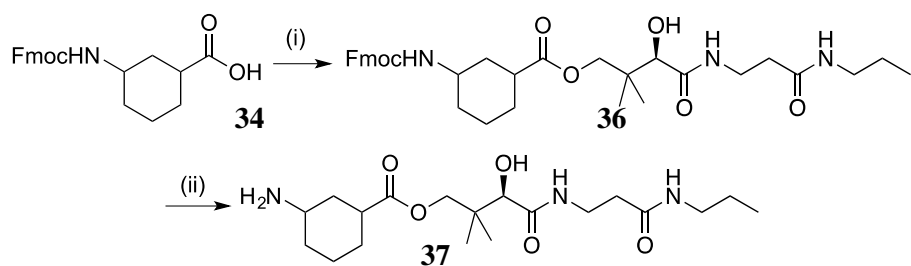
**Figure 3.13.** Synthesis of **34**. (i) FmocOSu (1.2 eq), NaHCO<sub>3</sub> (1.7 eq), water, THF, r.t, 30 h, 95%.

The synthesis of (1*R*,3*R*)-3-(((9*H*-fluoren-9-yl)methoxy)carbonyl)amino)cyclopentane-1-carboxylic acid, **35** from the commercially available *N*-Boc protected carboxylic acid was done in two steps, deprotection of the -Boc group using TFA in CH<sub>2</sub>Cl<sub>2</sub> followed by Fmoc protection using FmocOSu.

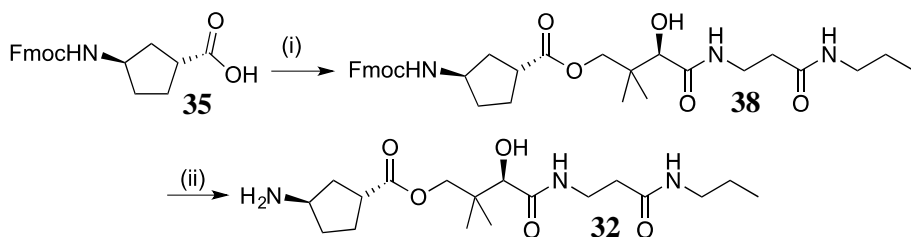


**Figure 3.14.** Synthesis of **35**. (i) TFA (xs), CH<sub>2</sub>Cl<sub>2</sub>, r.t, 15 h, quantitative; (ii) FmocOSu (1.3 eq), NaHCO<sub>3</sub> (1.7 eq), water, THF, r.t, 22 h, quantitative.

The Fmoc-protected linkers were then linked to the pantothenamide tail, **20** using DCC and DMAP, affording **36** in 37% and **38** in 32%. The compounds were stored in the Fmoc-protected forms and deprotection to free amine forms was performed when the coupling reaction to the heteroaromatic core was ready, because the free amines **37** and **39** (the removal of Fmoc group was confirmed quickly by <sup>1</sup>H-NMR) easily deteriorate.

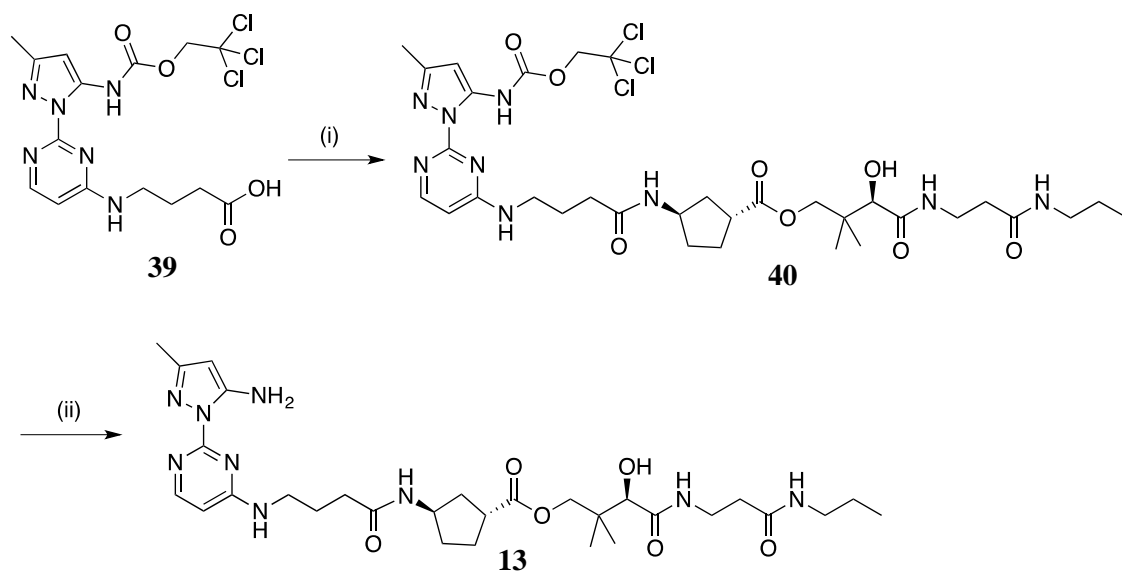


**Figure 3.15.** Synthesis of **37**. (i) DCC (2.6 eq), DMAP (0.6 eq), THF, 0°C, 1 h, **20** (1.4 eq), 45°C, 18 h, 37 %; (ii) piperidine (xs), THF, 4 h, 67%.

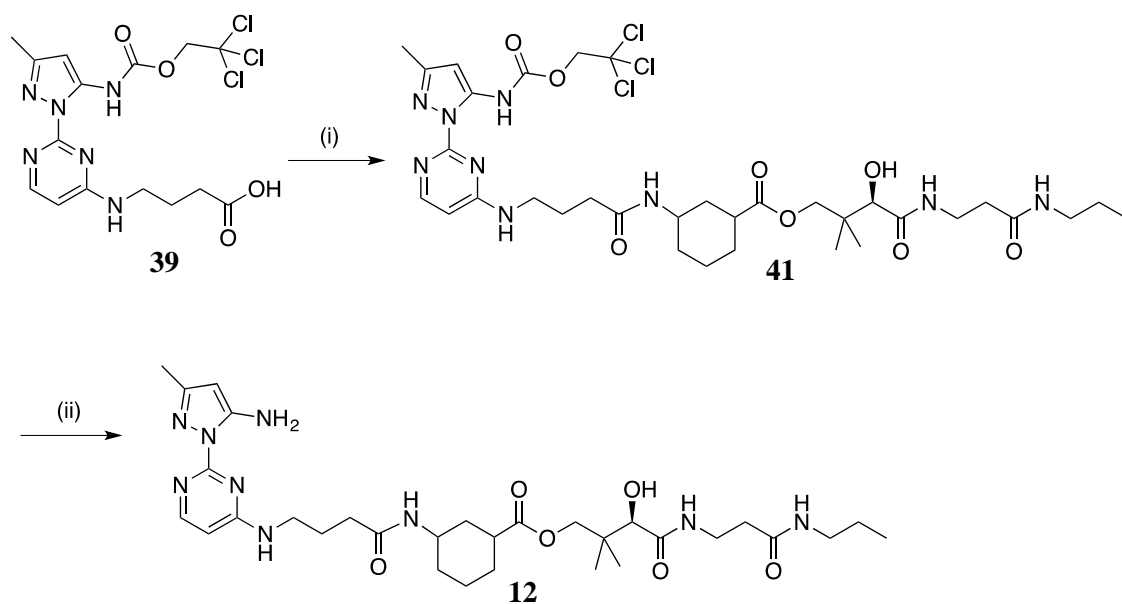


**Figure 3.16.** Synthesis of **32**. DCC (2.2 eq), DMAP (0.2 eq), THF, 0°C, 1 h, **20** (1.2 eq), 45°C, 24 h, 32 %; (ii) piperidine, THF, r.t, 3 h, 80%.

The final coupling reaction of **32** to the heteroaromatic core **39** (which was provided by my colleague, Fiona Bellany) was carried out using HOAt, PyAOP, DIPEA in DMF to achieve product **40** in 17%. However purification of the next deprotection step of the -Troc protecting group using Zn dust and NH<sub>4</sub>OAc was not successful due to contamination in HPLC column. Nevertheless the crude compound was screened for biological activity. The final coupling reaction of **37** to the heteroaromatic core was carried out by my colleague, Fiona Bellany. The reaction was not repeated due to inactivity of compound **12** (IC<sub>50</sub> >7 mM) and crude **13** (IC<sub>50</sub> >2 mM) and my re-assessment of the computational work resulted in a change of approach to this project. (Chapter 2)



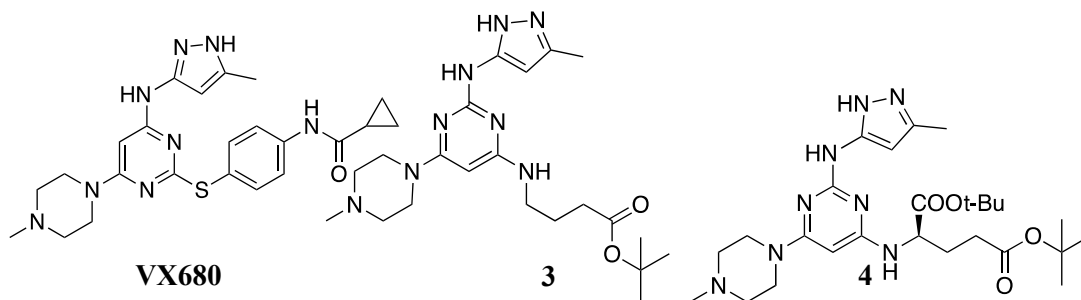
**Figure 3.17.** Attempted synthesis of **13**. (i) **32** (1.3 eq), HOAt (1.9 eq), PyAOP (3.6 eq), DIPEA, DMF, 0°C to r.t., 48 h, 17%; (ii) Zn, NH<sub>4</sub>OAc, THF : H<sub>2</sub>O = 8:1, r.t., 24 h, only crude obtained (<10%).



**Figure 3.18.** Synthesis of **12**. (i) **37**, HOAt, PyAOP, DIPEA, DMF, 0°C to r.t., 48 h, 22%; (ii) Zn, NH<sub>4</sub>OAc, THF : H<sub>2</sub>O = 8:1, r.t., 24 h, 78%.

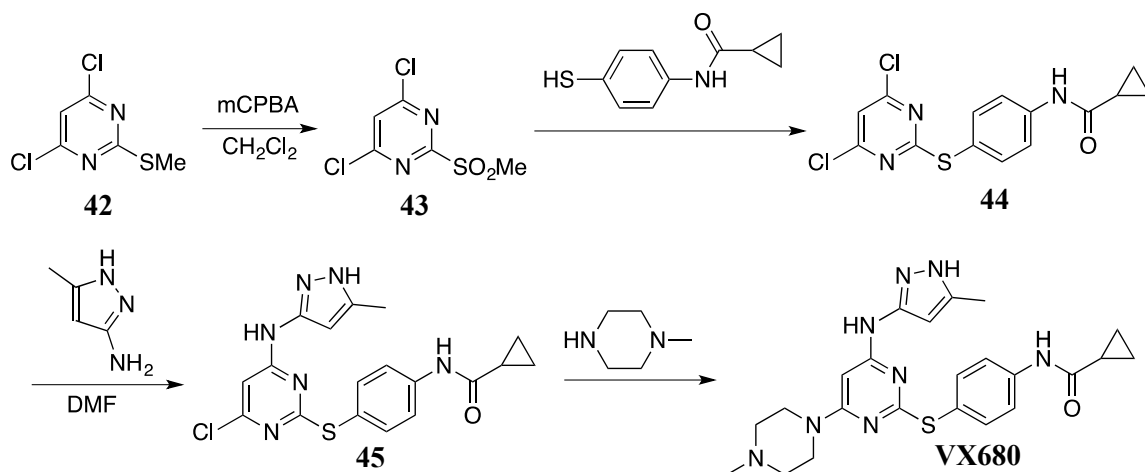
### 3.2. Heteroaromatic core

#### 3.2.1. Literature and preliminary research on synthesis of the heteroaromatic head group, the first generation of compounds



Based on the structure of VX680 and preliminary work by my colleague, Fiona Bellany (Table 1.3), compounds **3** and **4** were initially set as two of the target heterocycles. The regiochemistry of 3-amino 5-methylpyrazole substituent (C-2 of pyrimidine ring in **3** and **4**), which is different from VX680 (C-4 of pyrimidine ring), was set based on the main target **1** and **2**.

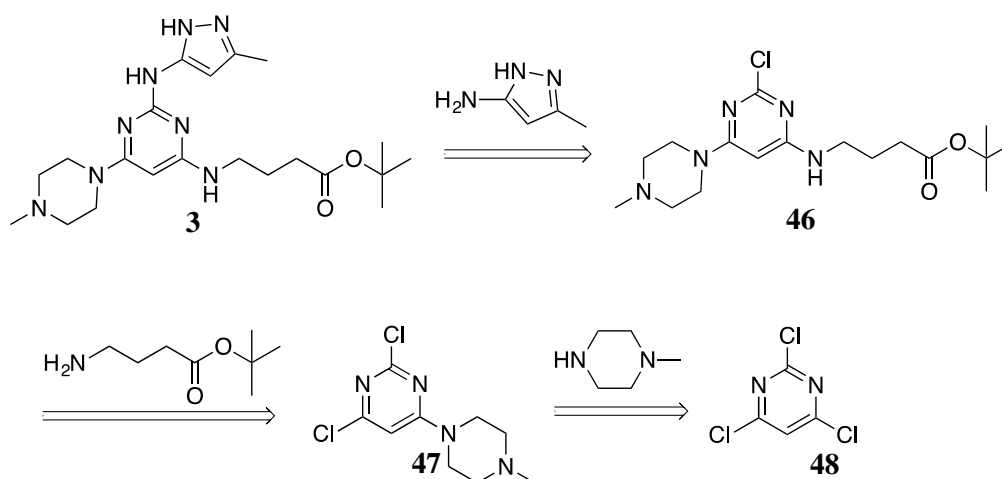
VX680 patented synthesis is shown in Figure 3.19.<sup>58b</sup> The –SMe at C-2 of pyrimidine ring was oxidized to a better leaving group –SO<sub>2</sub>Me. Nucleophilic substitutions allows formation of **44**. It was not discussed in the patent whether a competition was observed at C-2 and C4 positions. Similarly, nucleophilic substitution of 3-amino 5-methyl pyrazole to C-4 position was reported to occur at the primary amine of the pyrazole ring system. No competition between N-1 of pyrazole ring and –NH<sub>2</sub> at C-3 of pyrazole as nucleophiles was mentioned.



**Figure 3.19.** Patented synthesis of VX680 by Vertex Pharmaceuticals.<sup>58b</sup>

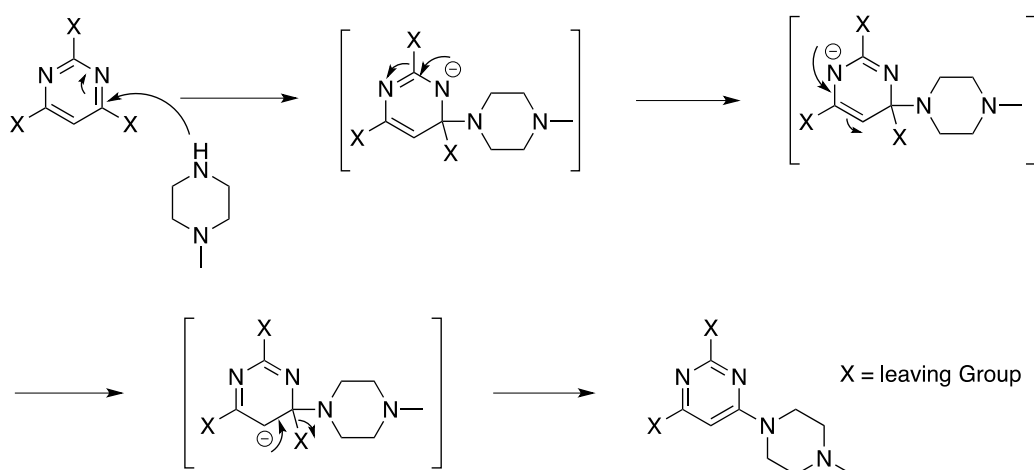
#### Retrosynthesis

The synthesis of the core structures was envisaged to involve a combination of S<sub>N</sub>Ar reaction and Buchwald-Hartwig reaction<sup>131</sup>. The retrosynthetic plan of **3** is shown in Figure 3.20. The 2,4,6-trichloropyrimidine was chosen as starting material due to its commercial availability. Synthesis of **4** would be similar as that of **3**.



**Figure 3.20.** Retrosynthesis of **3**

The mechanism of  $S_NAr$  reaction is shown in Figure 3.21 where the nucleophile (*N*-methyl piperazine) attacks the pyrimidine ring, forming the intermediate anions (Meisenheimer Complex). Pyrimidine readily undergoes  $S_NAr$  at the C-2 and C-4/6 positions with 4/6 being generally more reactive.<sup>132</sup> The reformation of the aromatic system and the leaving of chloride anion results in the product. As such electron-withdrawing substituents increase the reactivity of ring system as they stabilize the intermediate carbanions. (-Cl > -Br >> -I)

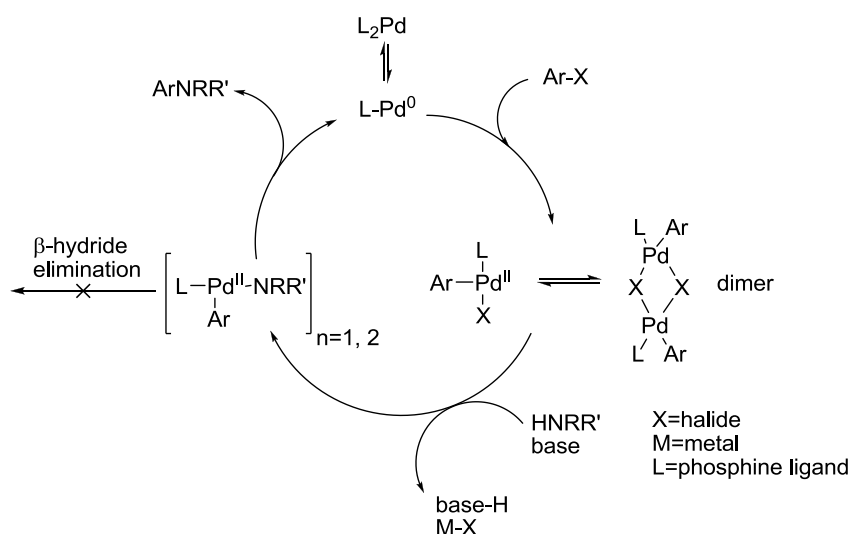


**Figure 3.21.**  $S_NAr$  reaction mechanism of reaction between *N*-methyl piperazine and trisubstituted pyrimidine.

In the case of weak nucleophiles where  $S_NAr$  reactions do not work, the assistance of a catalyst may be needed and a Buchwald-Hartwig reaction may be considered.

Reaction between the substituted pyrimidine ring and 3-amino 5-methyl pyrazole was predicted to be low yielding because of the possible competition between  $-NH_2$  substituent at C-3 of pyrazole ring and N-1 of pyrazole ring as nucleophiles. As a result, it was set as the final step of the synthesis. (This was also to make sure that no competitive substitutions between C-2 and C-4 positions of the pyrimidine ring needed to be considered). Consequently, a Buchwald-Hartwig reaction (palladium-catalysed coupling of heteroaryl amines and heteroaryl halides) was chosen for the reaction between the substituted pyrimidine ring and 3-amino 5-methyl pyrazole.

The proposed catalytic cycle of Buchwald-Hartwig coupling reaction from the literature <sup>133</sup> is shown in Figure 3.22.

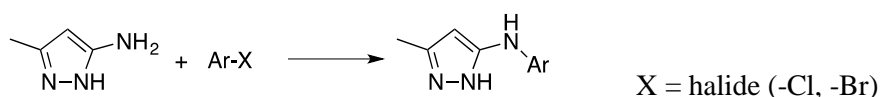


**Figure 3.22.** Proposed catalytic cycle of Buchwald-Hartwig coupling between amine and aryl halide. <sup>133</sup>

Buchwald *et al.* and Hartwig *et al.* <sup>133</sup> independently studied the coupling reactions between aryl halides and substituted amines using Pd catalysts, particularly the use of Pd(dba)<sub>2</sub> and phosphine ligands. The type of phosphine used is essential for the success of the coupling and the reaction is also sensitive to the type of base used. In the literature, <sup>134</sup> sterically hindered and electron-rich ligands such as DavePhos was reported to catalyse the reaction between aryl chloride and diethyl malonate. Literature research also suggests that (dba)Pd(BINAP) catalyst <sup>135</sup> may work particularly well with chloro-substituted pyrimidine systems. It is reported that under conditions of excess amine, amine will bind to Pd prior to oxidative addition of the aryl halide and (R(R')NH)Pd(BINAP) proceeds faster than does oxidative addition to Pd(BINAP). <sup>136</sup>

Most of the literature work reports palladium-catalysed coupling between heteroaryl amines and heteroaryl bromides or aryl bromides. It is suggested that due to bond length that may affect the initial rate of oxidative coupling reaction, aryl or heteroaryl bromides may react better than aryl or heteroaryl chloride. On the other hand, aryl and heteroaryl chlorides are generally more commercially available than their bromo analogues. Moss *et al.* reported that Pd-catalysed coupling reaction between 3-amino 5-methylpyrazole system and pyrimidine systems might occur at r.t in moderate to high yield. They also demonstrated that 2-chloropyridine reacted in a comparable time and in similar yield to 2-bromopyridine, concluding that their conditions might not be limited only to aryl bromides. (Table 3.1) <sup>137</sup>

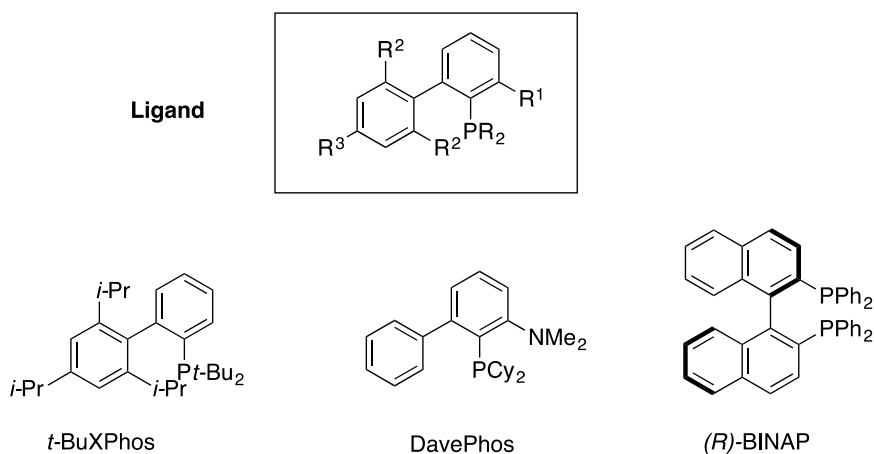
**Table 3.1 – Room temperature coupling of 3-amino-5-methylpyrazole with heteroaryl halides with *t*-BuXPhos precatalyst<sup>137</sup>**



Aryl Halide	Time (h)	Product	Yield %
	4		88
	6		89

Reaction conditions: 1.5-mmol scale, amine (1.0 eq), aryl halide (1.25 eq), NaO*t*-Bu (2.1 eq), *t*-BuXPhos precatalyst, L-Pd(0) (1 mol%) where L = *t*-BuXPhos ligand, *t*-BuXPhos ligand (1 mol%), *t*-BuOH (5 mL), r.t.

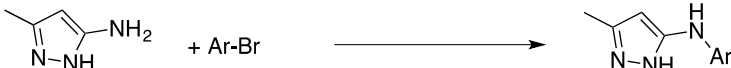
The conditions used in their report included *t*-BuXPhos ligand (a known effective catalyst in weakly nucleophilic heteroaryl system) however preliminary test by my colleague did not show promising results with this ligand. As a result, BINAP ligand was tested first before re-investigating *t*-BuXPhos ligand.

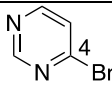
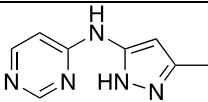
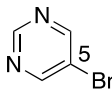
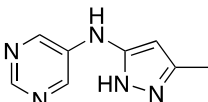
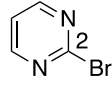
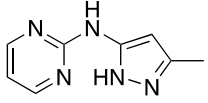


In terms of a Pd-catalysed coupling reaction between a 3-amino 5-methyl pyrazole system and a bromo pyrimidine system, Moss *et al.*<sup>137</sup> reported that no yield was isolated for cross coupling involving Br in 4/6-position of the pyrimidine ring. The reaction seemed to proceed best where Br was in the 5-position of the pyrimidine ring (yield of 92% isolated yield of 70%) followed by 2-position (isolated yield of 61%). (Table 3.2)



**Table 3.2 – Room temperature coupling of 3-amino-5-methylpyrazole with heteroaryl bromides with *t*-BuXPhos precatalyst<sup>137</sup>**



Aryl Bromide	Time (h)	Product	Yield %
	6		0
	4		(92) 70
	6		61

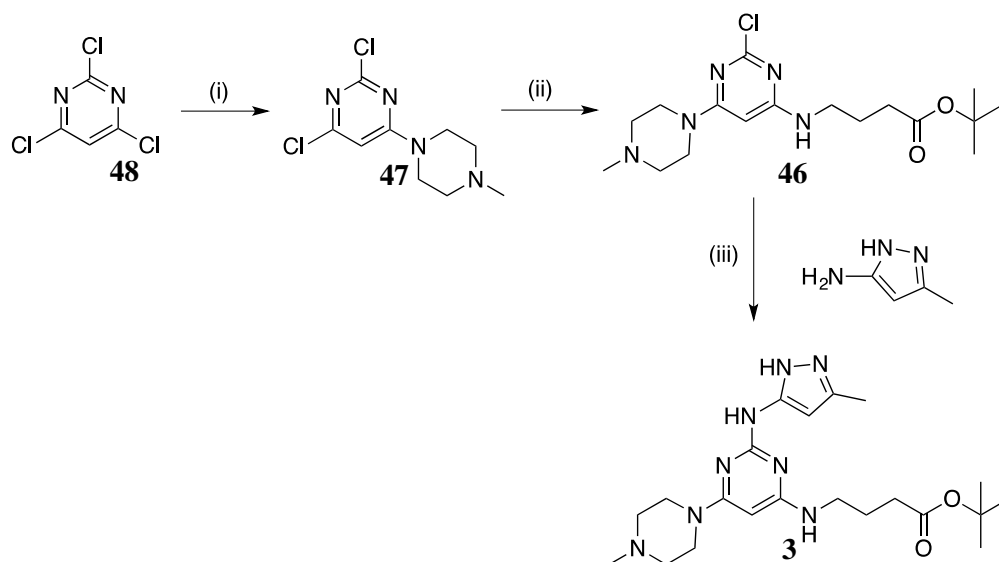
Reaction conditions: 1.5-mmol scale, amine (1.0 eq), aryl bromide (1.25 eq), NaO*t*-Bu (2.1 eq), *t*-BuXPhos precatalyst L-Pd(0) (1 mol%) where L = *t*-BuXPhos ligand, *t*-BuXPhos ligand (1 mol%), *t*-BuOH (5 mL), r.t.

### 3.2.2. Synthetic results

#### 3.2.2.1. Synthesis of 3

Reaction between 2,4,6-trichloropyrimidine and *N*-methyl piperazine proceeds well at -25°C to give the desired product 2,4-dichloro-6-(*N*-methylpiperazin-1-yl)pyrimidine, **47** in 74% yield.

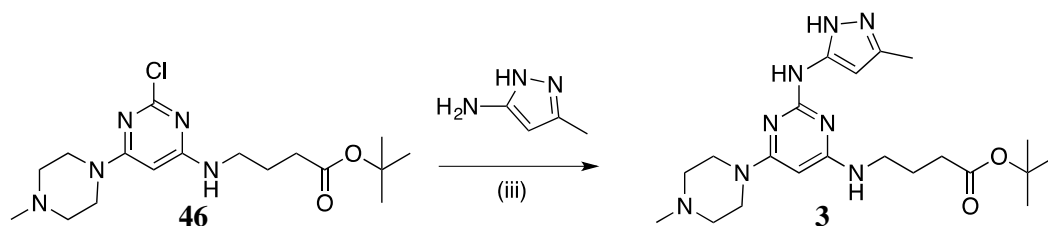
138



**Figure 3.23.** Synthesis of **3**. (i) *N*-methyl piperazine (1.0 eq), NEt<sub>3</sub> (1.1 eq), EtOH, -25 to 0°C, 4 h, 74%; (ii) *t*-butyl 4-aminobutanoate (1.8 eq), DIPEA (xs), EtOH, r.t., 26 h, 72%; (iii) 3-amino 5-methyl pyrazole (1.7 eq), (dba)Pd(*R*)-BINAP (15 mol%), Cs<sub>2</sub>CO<sub>3</sub>, THF, microwave 120 -140°C, 40 min, 80%.

*tert*-Butyl 4-aminobutanoate was added to **47** with DIPEA and EtOH under reflux conditions to give **46** in good yield (72%). The last coupling step with electron poor 3-amino-5-methyl pyrazole was carried out using Buchwald-Hartwig coupling. Several readily available ligands, solvents and bases (Table 3.3) were tried out. Refluxing proved unsuccessful but under microwave conditions the product was obtained in good yield with high mol% of catalyst (15 mol%). The total synthesis of core structure **3** is summarised above.

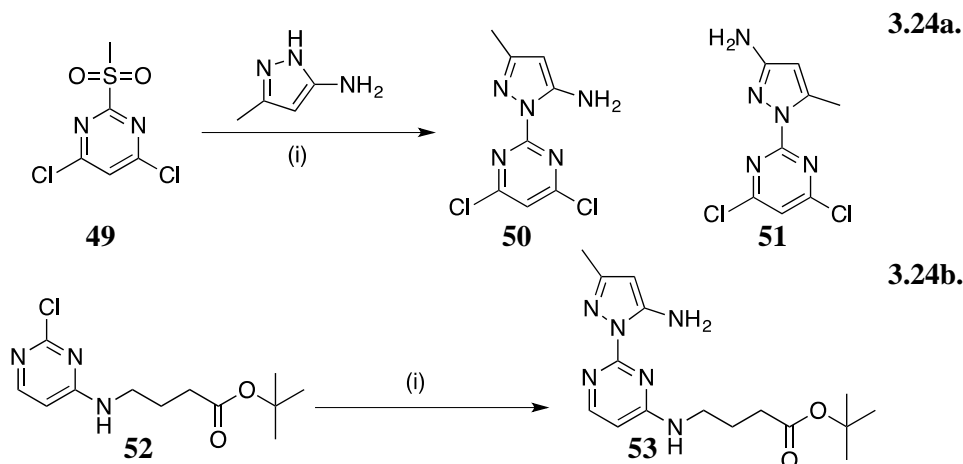
**Table 3.3**



(iii) Pd catalyst, ligand, base, solvent

Entry	Catalyst and ligand, mol%	Base	Solvent	Time (h)	Temperature	Yield %
1	(dba)Pd( <i>R</i> -BINAP), 1.3 <sup>137</sup>	Cs <sub>2</sub> CO <sub>3</sub>	dioxane	25	thermal heat, 100 °C	✗
2	Pd <sub>2</sub> (dba) <sub>3</sub> , DavePhos, 3.8 <sup>139</sup>	NaO <i>t</i> -Bu	dioxane	17	thermal heat, 100°C	✗
3	(dba)Pd( <i>R</i> -BINAP), 4	NaO <i>t</i> -Bu	dioxane	40	thermal heat, 95°C	✗
4	Pd(PPh <sub>3</sub> ) <sub>4</sub> , <i>R</i> -BINAP, 8	NaO <i>t</i> -Bu	toluene	27	thermal heat, reflux	✗
5	(dba)Pd( <i>R</i> -BINAP), 5	Cs <sub>2</sub> CO <sub>3</sub>	THF	1.6	microw, 140°C	6% isolated product
6	(dba)Pd( <i>R</i> -BINAP), 15	Cs <sub>2</sub> CO <sub>3</sub>	THF	0.6	microw, 120 – 140°C	80% isolated product

It was assumed that the primary amine would attach at the C-2 position of pyrimidine as the literature suggested. As the <sup>1</sup>H-NMR was obtained in CD<sub>3</sub>OD, the peaks of -NH and -NH<sub>2</sub> of the pyrazole could not be identified and structure **3** was therefore assumed to be correct. However, during the course of this research, the regioselectivity of similar reactions was investigated by X-ray crystallography by my colleague.

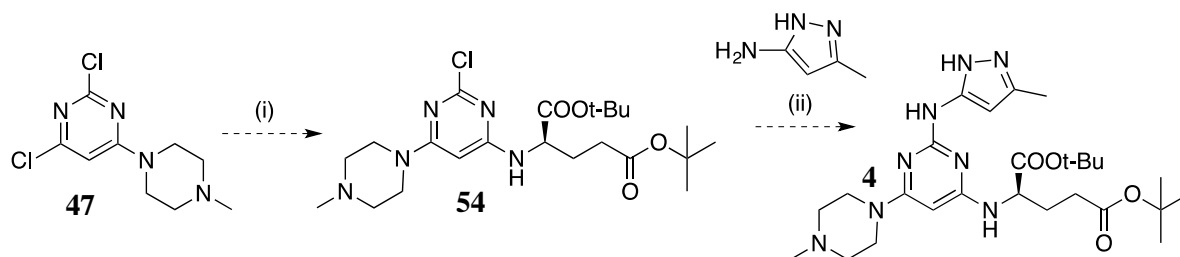


**Figure 3.24.** Investigation on regiochemistry of 2-chloro substituted pyrimidine system and 3-amino 5-methylpyrazole. a)  $S_NAr$  conditions: (i) NaH,  $CH_2Cl_2$ , reflux; b) Buchwald-Hartwig coupling conditions: (i) Pd (0), DavePhos, reflux.

Under strong base (NaH) conditions,  $S_NAr$  reaction showed that the major product, **50** was obtained through the attack of secondary amine of 3-amino 5-methylpyrazole. (Figure 3.24a) Further investigation reported that in neutral conditions the attack of the primary amine was favoured.<sup>140</sup> Crystal structure of the product **53** where Pd conditions with DavePhos as a ligand was not obtained. (Figure 3.24b) The two NMRs (in  $CDCl_3$ ) were compared by my colleague. They were found to be identical so reasonably it was assumed that in both case, the wrong regioisomers where the secondary amine was attached to C-2 of pyrimidine, were obtained. No further investigation was made for regiochemistry of **3** due to re-examination of target structures. (Section 1.5)

### 3.2.2.2. Progress towards syntheses of **4**

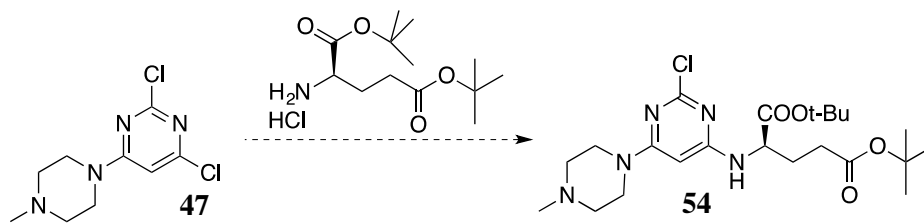
Synthesis of **4** was envisaged to follow the same route as the synthesis of **3**. (Figure 3.25)



**Figure 3.25.** Synthetic plan of core **4**.

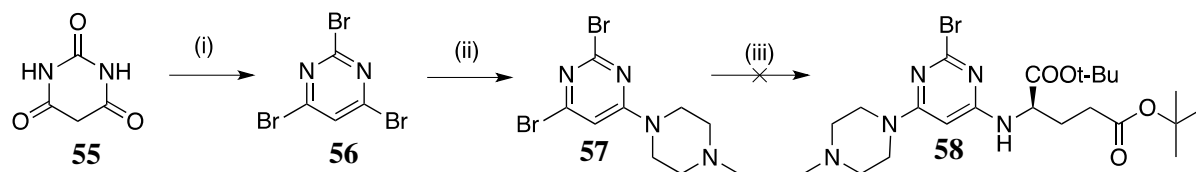
The first step is the synthesis of 2,4-dichloro-6-(*N*-methylpiperazin-1-yl)pyrimidine, **47**, as mentioned in synthesis of **3**. Synthesis of **54** was attempted using **47** and commercially available di-*tert*-butyl D-glutamate hydrochloride or free-form available di-*tert*-butyl D-glutamate but all had been unsuccessful. (Table 3.4)

**Table 3.4 – Attempts towards synthesis of 54**



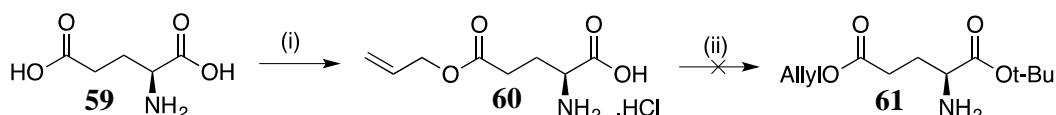
Entry	Amine (1.5-2 eq)	Base (+catalyst) (xs)	Solvent	Temperature	Time (h)	Yield %
1	salt	DIPEA	<i>n</i> -BuOH and EtOH	reflux	7	6% of crude <b>54</b>
2	salt	DIPEA	<i>n</i> -BuOH	reflux	40	X
3	salt	DIPEA	THF	reflux	4	X
4	salt	Et <sub>3</sub> N	EtOH	reflux	24	X
5	free	DIPEA	<i>t</i> -BuOH	reflux	24	X
6	free	DIPEA	<i>n</i> -BuOH and EtOH	reflux	72	X
7	salt	Cs <sub>2</sub> CO <sub>3</sub> (+(dba)Pd(( <i>R</i> )- BINAP))	toluene	microw, 140°C	25	X
8	salt	Cs <sub>2</sub> CO <sub>3</sub> (+(dba)Pd(( <i>R</i> )- BINAP))	THF	microw, 140°C	30	X

At the same time, 2,4,6-tribromopyrimidine was synthesised<sup>141</sup> as an alternative to 2,4,6-trichloropyrimidine. The bromo-substituted heterocycle was intended to be used in the case of difficult Buchwald-Hartwig reaction. Unfortunately the S<sub>N</sub>Ar reaction prior to the Buchwald-Hartwig coupling could not proceed. This synthetic attempt is described in Figure 3.26. Compound **56** was converted to 2,4-dibromo-6-(4-methylpiperazin-1-yl)pyrimidine **57** in 56% yield. Heating under reflux of **57** with di-*tert*-butyl D-glutamate hydrochloride in the presence of DIPEA as base and *n*-BuOH as solvent also failed to yield any sufficient product.



**Figure 3.26.** Synthetic attempt of **4** from 2,4,6-tribromopyrimidine. (i) POBr<sub>3</sub> (4.0 eq), PhNMe<sub>2</sub> (1.9 eq), toluene, reflux, 3 h, 68%; (ii) *N*-methyl piperazine (1.0 eq), NEt<sub>3</sub> (2.3 eq), EtOH, -30 to -5°C, 3 h 56%; (iii) di-*tert*-butyl D-glutamate hydrochloride (1.2 eq), DIPEA (1.2 eq), *n*-BuOH, 45°C, 2 h then reflux, 15 h.

The reason for the failure of the reaction between **47** and di-*tert*-butyl D-glutamate hydrochloride was postulated that steric hindrance from one or both *t*-Bu esters prevented the reaction from taking place. As selective deprotection at the  $\delta$ -COOH would be required for later coupling to the linker and pantothenamide, the synthesis of Glu(OAllyl)Ot-Bu, **60**, was attempted. Trials were carried out on L-glutamic acid. Allyl protection of glutamic acid proceeded well in moderate yield, 42%.<sup>142</sup> (Figure 3.27) However, later *tert*-butyl protection using perchloric acid (xs) and *t*-butyl alcohol as solvent under stirring condition at r.t for 72 h was not successful.



**Figure 3.27.** Di-protection of L-Glutamic acid. (i) Allyl alcohol (xs), TMSCl (xs), 0°C to r.t, 26 h, 42%; (ii) perchloric acid (8.3 eq), *t*-butyl alcohol (13.0 eq), r.t, 48 h.

Another suggestion to be tested was to change the leaving group (-Cl, -Br) at C-6 or C-2 of pyrimidine to a better leaving group such as a tertiary amine salt or fluoride before attempting other nucleophilic substitutions.

### 3.3. Conclusion

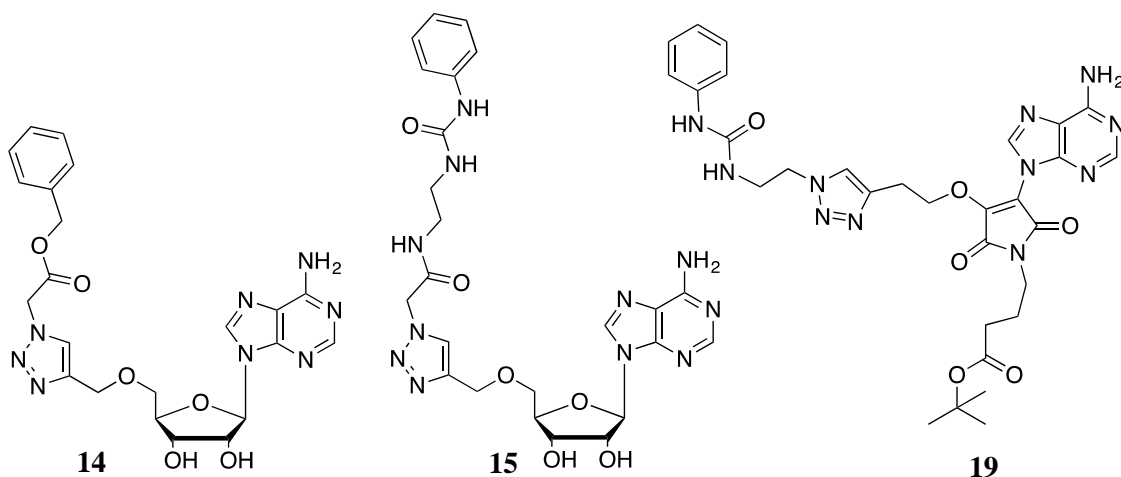
The main difficulty of Approach A synthesis lies in the heterocyclic chemistry, in particular the reaction between 3-amino-5-methyl pyrazole and 2,4,6-substituted pyrimidine. The heterocyclic chemistry was investigated in detail by my colleague. Unfortunately, revised computational docking showed that the binding mode of VX680-like head group changed completely with the addition of pantothenamide end. This disrupted the research on Approach A synthesis. The chemistry of the second generation of compounds is described in Chapter 4.

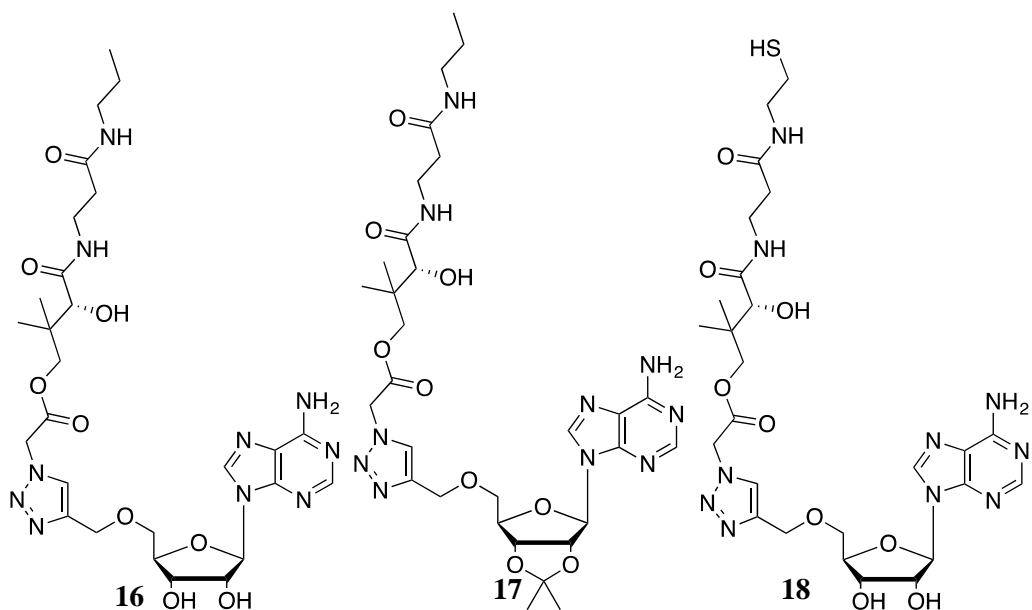
#### 4. Approach B syntheses

In this chapter, work has been focused on using the triazole group as an artificial chemical linkage to replace the pyrophosphate link in CoA. As outlined in Chapter 1 (Section 1.5.2) and Chapter 2 (Section 2.4.2) this linker has been chosen for further study for four reasons:

- (i) The feasibility of forming the prerequisite alkynes and azides that are attached to pantothenamide tail and head group.
- (ii) The possibility of the desired stable products to be formed in aqueous media (from functional groups that may be orthogonal to those in CoA), under mild conditions with sufficient yield.
- (iii) The triazole ring system is chemically robust and non-toxic and it has a limited number of rotatable bonds, which would provide more accurate prediction in modeling.
- (iv) Allowing a robust synthesis to provide a set of compounds, containing a diverse range of functional groups and length, in a short period of time.

With the aim to establish an understanding of the ATP binding pocket of AURKA, including the activation loop which is disordered, computational study determined 6 main compounds to synthesise (as mentioned in Figure 2.39):



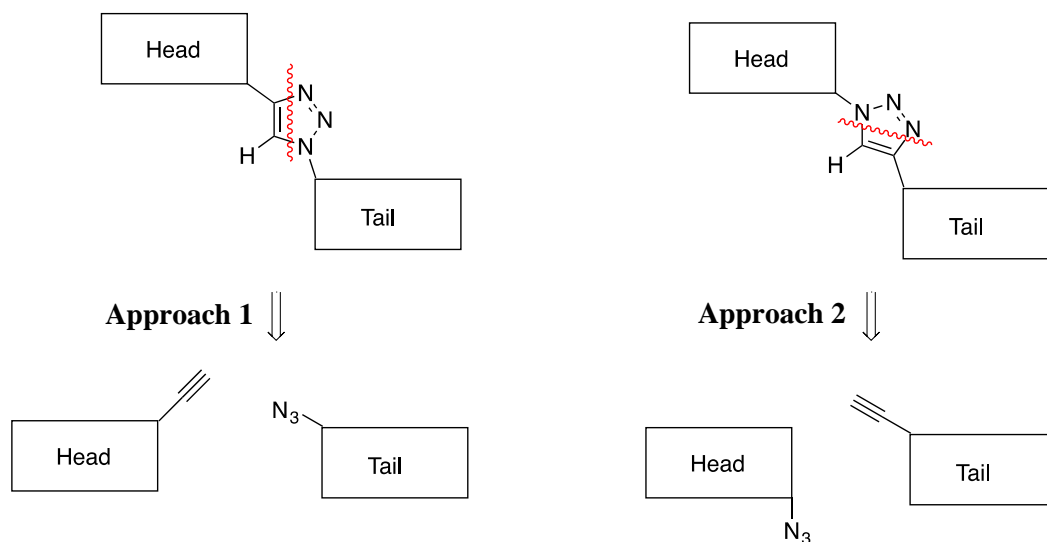


#### 4.1. Literature research and proposed approach

*The feasibility of forming the prerequisite alkynes and azides*

Retrosynthetic analysis suggested two approaches to prepare these analogues: Approach 1: Head-alkyne, X and Tail-azide and Approach 2: Head-azide, X and Tail-alkyne. (Figure 4.1)

There were two aspects to be investigated: one was to test whether it is possible to prepare prerequisite alkynes and azides and the other was to test the feasibility of Click reaction between prerequisite alkynes and azides.

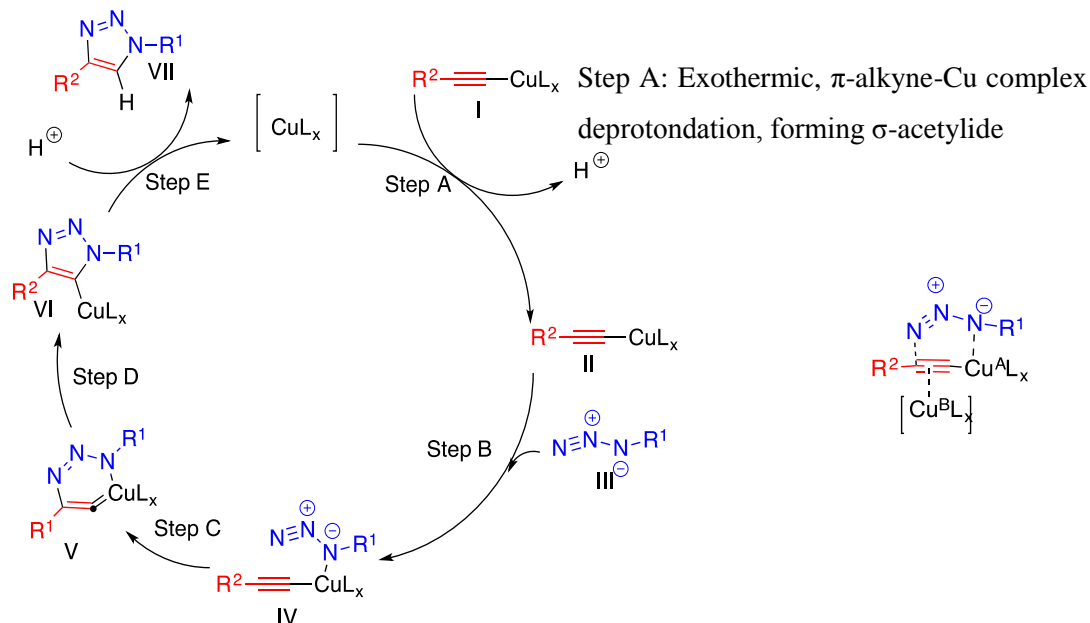


**Figure 4.1.** Retrosynthesis of “triazole-linked compounds”

Because the pantothenamide tail was only able to withstand mild reacting conditions (Section 3.1), it was preferred that tail-azide/tail-alkyne is formed first (lower yield reaction) before the Cu (I)-catalysed alkyne-azide coupling (CuAAC) reaction. However if this route proves difficult, the other approach is to perform CuAAC reaction first, followed by esterification reaction to link with the pantothenamide tail.

*The feasibility of Click reaction between the required alkyne and azide*

The general proposed reaction mechanism of a CuAAC is presented in Figure 4.2. Cu (I) catalysed CuAAC reaction was reported to produce a 1,4-substituted triazole system.<sup>143</sup>



**Figure 4.2.** General mechanism of CuAAC reactions.<sup>143</sup>

It is notable that CuAAC reactions, particularly with biomimetic systems may take 24 to 48 hours to proceed with significant yield. For the purpose of generating later a large library of analogues, there are 5 features to achieve when investigating this type of reaction. Reactions should proceed in aqueous media, under mild and simple conditions (preferably not oxygen-free condition), tolerable to different types of substrates and are able to generate significant yield with easy purification procedure.

In order to achieve these features, there were two features from the literature to highlight before investigating the feasibility of CuAAC reaction in this case and its optimized conditions. The first one is **catalysts**. The common source of copper (I) catalyst is  $CuSO_4$ , which can be reduced *in situ* by ascorbate to  $Cu^I$ .<sup>144</sup> Copper wire ( $Cu^0$ ) usually resulted in longer reaction time at r.t and was used for microwave-assisted Click reaction.<sup>145</sup> Preparation of reactive Cu (I) acetylide *in situ* (in the presence of an organic azide) is critical for the success. Formation of polymeric acetylide is detrimental for the reaction hence to be avoided. As a result, stabilising ligands e.g. TBTA and THPTA were used to prevent the formation of unreactive polynuclear Cu (I) acetylide and facilitate the coordination of the azide to Cu centre and increase the solubility of Cu complex.<sup>146</sup> The second highlighted feature is **solvents**. Water is an ideal solvent capable of supporting reactive Cu (I) acetylide.



## 4.2. Synthetic results

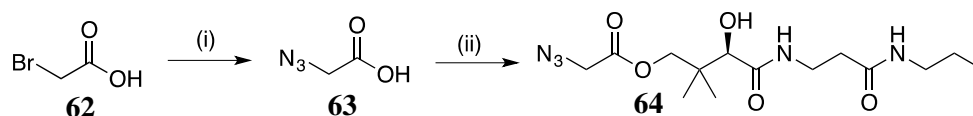
### 4.2.1. The feasibility of forming prerequisite alkynes and azides

Out of the two approaches, Approach 1 proved to be more reliable and easier to follow, mainly due to the observation that the formation of tail-azides is more feasible than that of tail-alkynes.

#### *Approach 1 results*

##### *“Tail-azide” synthesis*

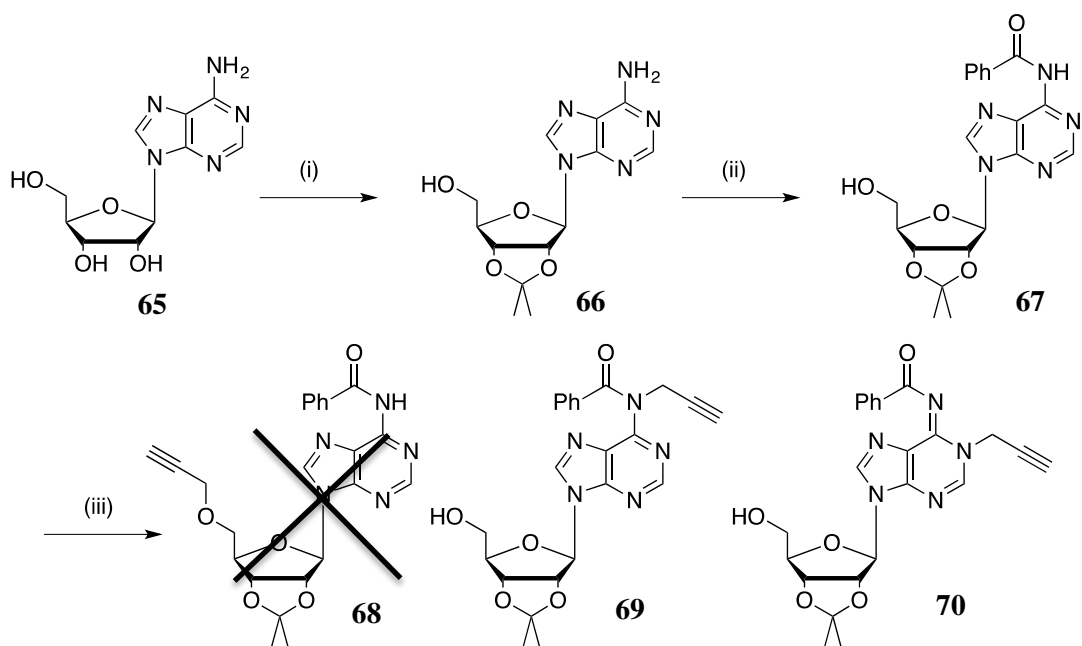
The pantothenamide-linked azide was prepared from an esterification between pantothenamide, **20** and 2-azidoacetic acid, **63** using DCC, DMAP or PyAOP, HOAt as coupling reagents. DCC, DMAP was chosen as the preferred method for economic reason although the reaction yield was lower than that using PyAOP and HOAt. 2-Azidoacetic acid, **63** was prepared using a method from literature procedure<sup>147</sup> from commercially available 2-bromo acetic acid, **62** in quantitative yield. (Figure 4.3)



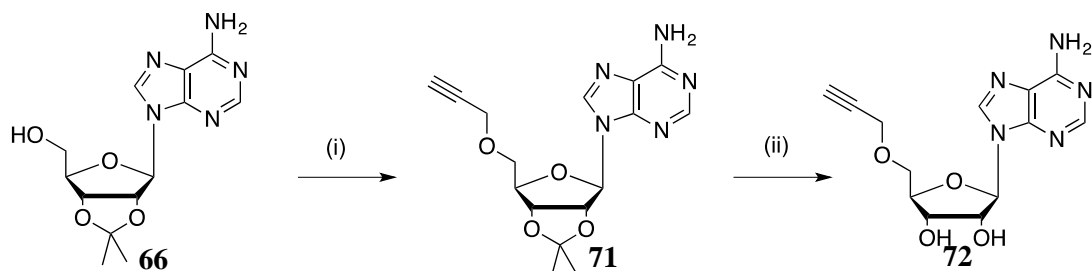
**Figure 4.3.** Synthesis of **64**. (i) Sodium azide (xs), water, quantitative; (ii) **20** (0.7 eq), DCC (1.2 eq), DMAP (0.1 eq), THF, 0°C to r.t., 20 h, 44%; or **20** (1.0 eq), PyAOP (1.9 eq), HOAt (1.5 eq), DIPEA (1.1 eq), DMF, 0°C to r.t., 18 h, 57%.

##### *“Head-alkyne” synthesis*

In order to prepare **72** (Figure 4.5), the prerequisite alkyne for many of the targets, the route previously reported in the literature<sup>148, 149</sup> was first attempted. Ketalisation of the 2' and 3' –OH groups of commercially available adenosine<sup>148</sup> was followed by benzoyl protection of the adenine –NH<sub>2</sub> group<sup>149</sup>. However, subsequent attempts to selectively introduce the propargyl group under literature conditions<sup>149</sup> using sodium hydride, propargyl bromide and TBAI (phase transfer reagent) in THF to produce **68** failed. Instead, unwanted products, **69** and **70**, ratio 1:2, were found in the product mixture. (Figure 4.4) As a result, non-protective strategy was carried out. This proved to be successful, giving **72** in similar yield to that reported in the literature in less synthetic steps. (Figure 4.5)



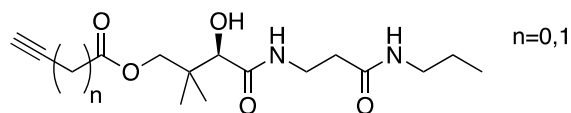
**Figure 4.4.** Attempted synthesis of **68**. (i) *p*-toluenesulfonic acid (xs), acetone, 0°C to 11°C, 20 h, quantitative; (ii) TMSCl (xs), BzCl (1.3 eq), pyridine, r.t, 3.5 h then water, NH<sub>3</sub> (xs), 0°C, 10 min, 64%; (iii) NaH (1.1 eq), THF, 0°C, 2 h, then TBAI (0.3 eq), propargyl bromide (1.3 eq), 0°C to r.t, 20 h.



**Figure 4.5.** Synthesis of **72**. (i) NaH (1.5 eq), propargyl bromide (1.5 eq), THF, 0°C to r.t, 24 h, 41%; (ii) TFA (xs), *t*-BuOH, water, r.t, 24 h, quantitative.

### Approach 2 results

#### “Tail-alkyne”

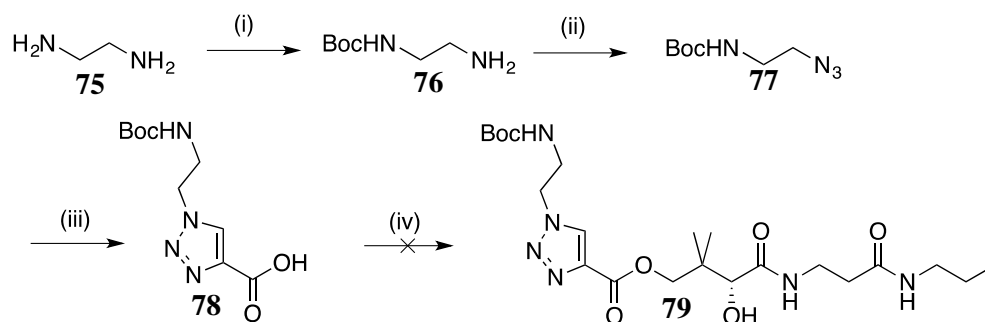


For compound X,  $n=1$ , proposed retrosynthesis starts with but-3-ynoic acid and pantothenic acid (free form) as starting materials, coupled together by an esterification reaction. However synthesis of but-3-ynoic acid from its alcohol via oxidation using NaIO<sub>4</sub>, K<sub>2</sub>Cr<sub>2</sub>O<sub>7</sub>, HNO<sub>3</sub><sup>150</sup> was not successful. (Table 4.1) As it is the very first step of the long synthesis and the reaction did not clearly provide high yield and easy purification, this was not pursued further.

**Table 4.1 – Tested oxidising conditions for but-3-yn-1-ol, 74**

Conditions	Results
NaIO <sub>4</sub> , K <sub>2</sub> Cr <sub>2</sub> O <sub>7</sub> , HNO <sub>3</sub> , water, r.t. <sup>150</sup>	polymerised

Propiolic acid was readily available for purchase so it was used as an alternative to test esterification reaction in this case. However, attempts to push the reaction forward using both mild conditions (coupling reagents) and forcing conditions (acyl chloride), keeping in mind the sensitivity of pantothenamide tail were unsuccessful. Literature suggested that esters of propiolic acid required forcing conditions (reflux in benzene for 2 - 16+ hours with *p*-toluenesulfonic acid)<sup>151</sup>. The other option would be to perform “Click chemistry” first, forming the triazole ring system before coupling it to the pantothenamide tail. A simple triazole ring system, **78** was synthesized to test the idea. (Figure 4.6) The “Click” conditions adapted from the literature<sup>152</sup> allowed **78** to be formed in good yield. (Table 4.2) However, the following esterification to link the pantothenamide, **20** with **78** were unsuccessful.



**Figure 4.6.** Attempted synthesis of **79**. (i) Boc<sub>2</sub>O (1.0 eq), CH<sub>2</sub>Cl<sub>2</sub>, r.t, 24 h, quantitative; (ii) NaN<sub>3</sub> (xs), (F<sub>3</sub>CSO<sub>2</sub>)<sub>2</sub>O (2.0 eq), CH<sub>2</sub>Cl<sub>2</sub>, water at 0°C, 2 h then work-up followed by **76**, MeOH, K<sub>2</sub>CO<sub>3</sub> (1.6 eq), CuSO<sub>4</sub>·5H<sub>2</sub>O (6 mol%), H<sub>2</sub>O, r.t, 24 h, 81%; (iii) propiolic acid (0.13 M, 1.1 eq), CuSO<sub>4</sub>·5H<sub>2</sub>O (15 mol%), sodium ascorbate (68 mol%), *t*-BuOH : water = 1:2, r.t, 48 h, 75%<sup>153</sup>; (iv) **20** (1.0 eq), PyAOP (4.0 eq), HOAt (0.5 eq), DIPEA (xs), 0°C to 60°C, 18 h.

**Table 4.2 – Testing “Click” conditions for preparation of 78**

[Reactant] M		Solvent ratio (A:B)		Catalyst loading (mol%)		Time	Yield
		<i>t</i> -BuOH	water	CuSO <sub>4</sub> ·5H <sub>2</sub> O	sodium ascorbate	stirring at r.t (h)	%
0.05	0.05	1	2	22	90	18 h	41
0.12	0.13	1	2	5.6	24	48 h	75

#### 4.2.2. The feasibility of Click reaction between alkyne **72** and azide **64**

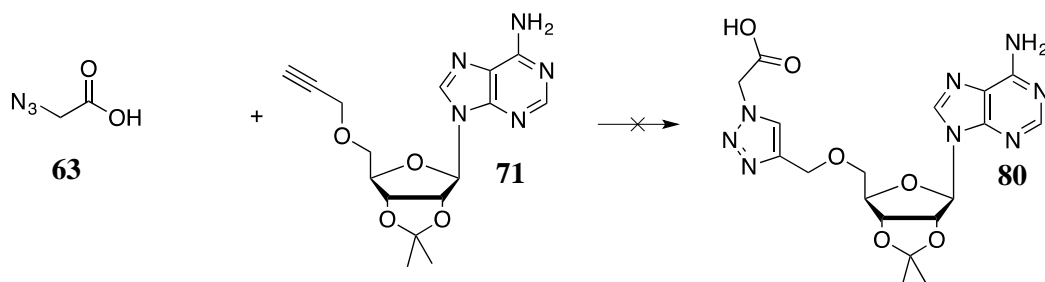
Having successfully synthesised alkyne **72** and azide **64**, investigation of CuAAC reaction between the two was carried out.

Several attempts to couple alkyne **71** and azide **64** using similar conditions for synthesizing **78** above were not successful. It was hypothesized that the concentration of the reactants and the ratio of them to catalyst loading might be crucial to the success. (Table 4.3)

**Table 4.3 – Initial tests for optimal Click conditions on alkyne and azide, 71 and 64**

Entry	[Reactant] M		Solvent ratio (A:B)		Catalyst loading (mol%)		Time stirring at r.t (h)	Yield %
	<b>64</b>	<b>71</b>	<i>t</i> -BuOH	water	CuSO <sub>4</sub> ·5H <sub>2</sub> O	sodium ascorbate		
1	0.1	0.06	1	1	4.8	20	48 h	X
2	0.02	0.02	1	2	6.5	23.5	48 h	X

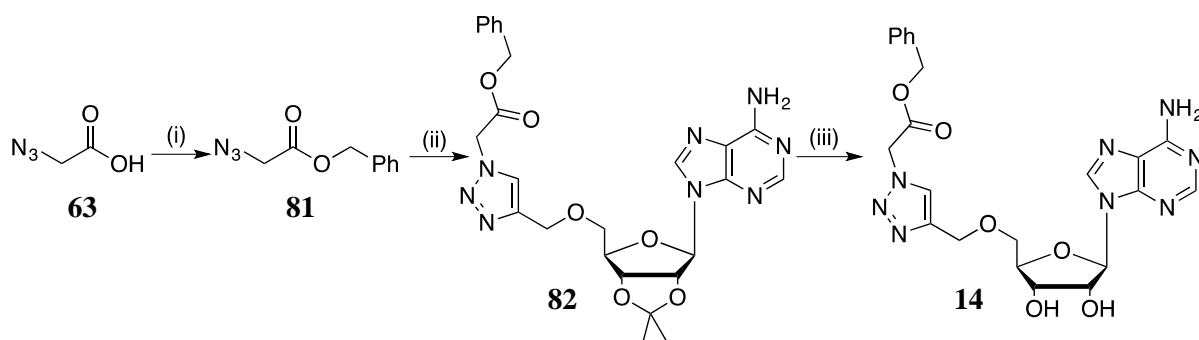
Because the literature review<sup>154, 155</sup> suggested that it was difficult to perform to completion the CuAAC reaction of adenosine-containing compounds in particular and of biomimetic systems in general (long reaction time and heating was required), a series of test experiments was carried out. Firstly, CuAAC test reaction was carried out between 2-azidoacetic acid **63** and **71** but no product was obtained. (Table 4.4)

**Table 4.4 – Second trials for optimal Click conditions**

Entry	[Reactant] M		Solvent ratio (A:B)		Catalyst loading <sup>†</sup> (mol%)		Time (h), T (°C)	Yield %
	<b>63</b>	<b>71</b>	<i>t</i> -BuOH	water	CuSO <sub>4</sub> ·5H <sub>2</sub> O	sodium ascorbate		
1	0.06	0.05	1	1	1.8	9	48 h, r.t	X
2	0.07	0.07	1	1	0.9	9	48 h, r.t	X
3	0.06	0.05	1	1	0.9	18	48 h, r.t	X
4	0.02	0.02	1	1	1.5	7.5	1 h, microw 100°C	X

<sup>†</sup>Catalysts' stock solution in water: [CuSO<sub>4</sub>·5H<sub>2</sub>O] = [sodium ascorbate] = 0.1 M for the first 3 entries and of 0.05 M for the last entry.

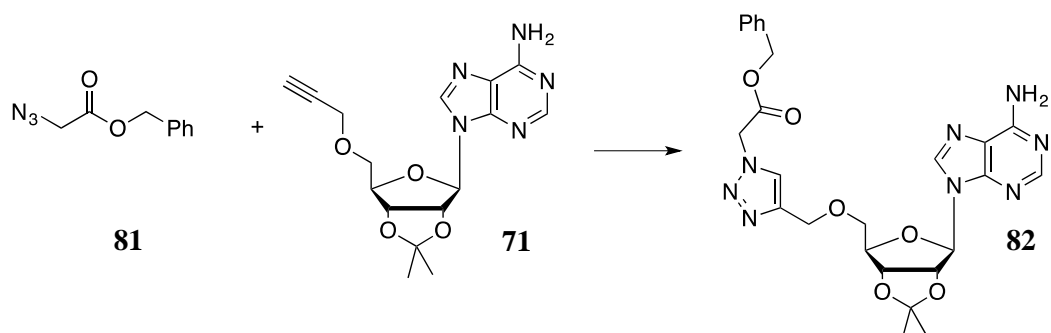
The failure was thought to be due to the fact that 2-azidoacetic acid was soluble in water, much more than **71**. This might cause difficulty to form complex **IV** (Figure 4.2). As such 2-azidoacetic acid was converted into benzyl 2-azidoacetate,<sup>156</sup> which was then successfully coupled to **71** to form **82**, the precursor of target compound **14**. Ketal removal of **82** by TFA afforded **14** in good yield, 82%. (Figure 4.7) The choice of a benzyl group was suggested by modeling work that showed a possible  $\pi$ -stacking interaction of a phenyl ring of a ligand with Phe144 in the ATP binding pocket of AURKA. Besides, the benzyl protecting group is known to be removed by H<sub>2</sub>, Pd/C (mild conditions) if required. Catalyst loading was low in these trials and might be increased. The trials to obtain optimised conditions for CuAAC reaction between azide **81** and alkyne **71** were shown in Table 4.5 and 4.6.



**Figure 4.7.** Synthesis of **14**. (i)  $\text{Cs}_2\text{CO}_3$  (0.5 eq), MeOH,  $\text{H}_2\text{O}$ , -3 to  $0^\circ\text{C}$ , 2 h then  $\text{PhCH}_2\text{Br}$  (1.0 eq), DMF,  $0^\circ\text{C}$  to r.t., 15 h, 72%<sup>95</sup>; (ii) **71** (1.0 eq),  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ , sodium ascorbate,  $t\text{-BuOH} : \text{H}_2\text{O} = 1:1$ , r.t., 24 h, 83%; (iii) TFA (xs),  $t\text{-BuOH} : \text{H}_2\text{O} = 1:1$ , r.t., 14 h, 82%.

Table 4.5 presents CuAAC third trials between benzyl 2-azidoacetate, **81** and **71**. Entry 1 proved to be successful after two weeks stirring at r.t with a yield of 59%. However, the reaction did not go to completion. In entry 6, 61% of **71** was reacted, giving an isolated yield of 56%. Entry 6 also suggested that for 0.02 M concentration of the reactants, even a 0.2 mol% of  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$  would catalyse the coupling. This indicates that the reaction proceeds in the below condition at slow rate. A slight increase in the concentration of reactants may speed up the reaction. In entry 5, when the amount of azide was initially a third of that of the alkyne for the first 24 h, reaction could not proceed further to achieve similar yield as that in entry 1. The suggested explanation was that polymeric Cu (I) acetylide might have formed and hindered the progress of the reaction.

**Table 4.5 – Third trials for optimal Click conditions**



Entry	[Reactant] M		Solvent ratio (A:B)		Catalyst loading † (mol%)		Time (h), T (°C)	Yield %
	<b>81</b>	<b>71</b>	<i>t</i> -BuOH	water	CuSO <sub>4</sub> ·5H <sub>2</sub> O	sodium ascorbate		
1	0.02	0.02	1	1	0.75	7.5	312 h, r.t	59
2	0.02	0.02	1	1	0.75	15	20 min, 100°C microw	Trace of product in LCMS
3	0.02	0.02	1	1	0.75	15	1.3 h, 100°C microw	Trace of product in LCMS
4	0.02	0.02	1	1	0.75	15	1.3 h, 100°C, then 1 h, 150°C microw	Degraded X
5	7 x 10 <sup>-3</sup> then 0.02	0.02	1	1	0.8	7.3	264 h, r.t ‡	25
6	0.02	0.02	1	1	0.2	1.85	192 h, r.t	56

† Catalysts' stock solution in water: [CuSO<sub>4</sub>·5H<sub>2</sub>O] = [sodium ascorbate] = 0.05 M. ‡ 24 h, r.t at 7 x 10<sup>-3</sup> M of **81** then 240 h at 0.02 M of **81**.

In CuAAC fourth trials, the concentrations of the reactants were varied based on the above results yielded the optimised conditions for CuAAC reaction involved adenosine-alkyne (Table 4.6).

**Table 4.6 – Fourth trials for optimal Click conditions**

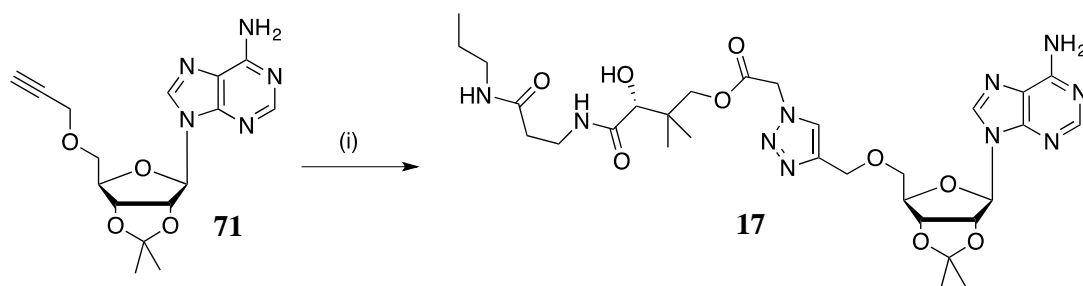
Entry	[Reactant] M		Solvent ratio (A:B)		Catalyst loading † (mol%)		Time (h), r.t	Completion %
	<b>81</b>	<b>71</b>	<i>t</i> -BuOH	water	CuSO <sub>4</sub> ·5H <sub>2</sub> O	sodium ascorbate		
1	0.06	0.06	1	1	0.8	7.5	26 h	90
2	0.15	0.15	1	1	0.8	7.5	26 h	80
3	0.06	0.06	1	1	0.8	7.5	24 h	96

†Catalysts' stock solution: [CuSO<sub>4</sub>·5H<sub>2</sub>O] = [sodium ascorbate] = 0.05 M.

Monitoring the reactions by LCMS showed that at the same catalyst loading, solvent system and other conditions, 0.06 molar concentration of reactants provided a faster rate of reaction (formation of product : consumption of reactants) of 1.5 times compared to 0.15 molar concentration of reactants. A scale-up of the reaction in entry 1 to six times in entry 3 showed the same observations.

In summary, the optimized conditions for CuAAC reaction involving adenosine-alkyne required the equimolar concentrations of azide and alkyne at 0.05 M or 0.06 M. The source of copper (I) was CuSO<sub>4</sub>·5H<sub>2</sub>O at 0.8 mol% and that of reducing agent was sodium ascorbate at 8 mol%, both from freshly made 0.05 M solutions. In case the coupling proved to be difficult, stabilising ligand, THPTA was used at 0.8 mol%. The temperature was r.t (20°C) and the solvent system was *t*-BuOH and H<sub>2</sub>O at ratio 1:1 or 1:2 respectively.

The optimised conditions were then used for the coupling between alkyne **71** and pantothenamide azide **64** to form target compound **17**. After 15 h, LCMS monitoring suggested that the reaction was 60% completed. Purification by flash column chromatography afforded product **17** in 56% isolated yield.

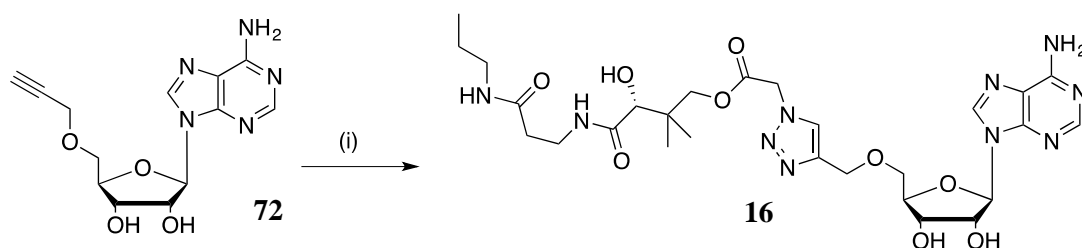


**Figure 4.8.** Synthesis of **17**. (i) **64** (1.0 eq), CuSO<sub>4</sub>·5H<sub>2</sub>O, sodium ascorbate, *t*-BuOH : H<sub>2</sub>O = 1:1, r.t, 15 h, 56%.

In case of **16**, in the absence of a stabilising ligand, the reactions appeared to be slow (not up to 80-90% completion after 24 h). Therefore THPTA was added as the stabilising agent. This allowed the reactions to be completed within 1 - 2 h. (Figure 4.9) The disadvantage was that it requires extremely inert conditions (O<sub>2</sub>-free) to avoid re-oxidation of Cu (I) to Cu (II). One solution to this was to keep adding reducing agent (sodium ascorbate) to the reaction mixture or copper wire to promote disproportionation reaction to keep providing Cu (I) source for the

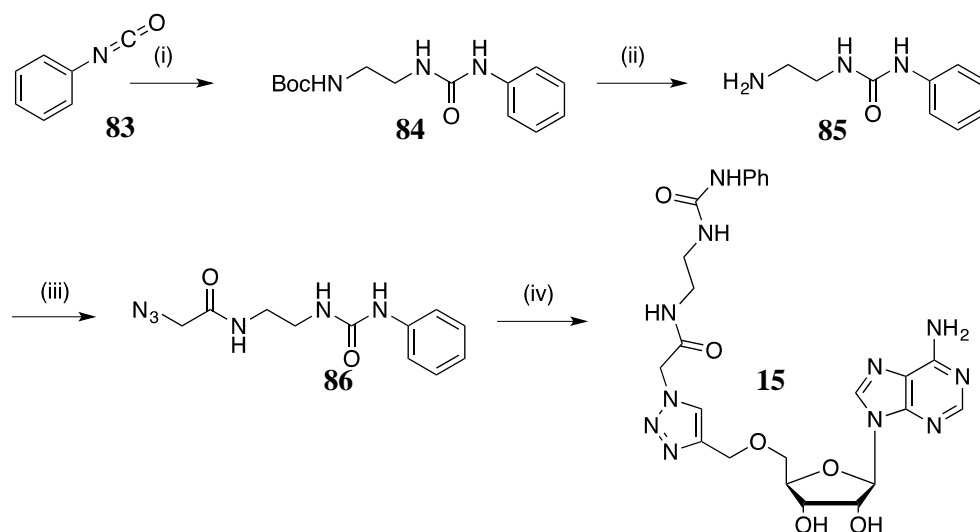


reaction. During these reactions, sodium ascorbate was chosen to be added in the case where re-oxidation of Cu (I) to Cu (II) occurred, as the reaction time was short.



**Figure 4.9.** Synthesis of **16**. (i) **64** (1.0 eq), CuSO<sub>4</sub>·5H<sub>2</sub>O, THPTA, sodium ascorbate, *t*-BuOH : H<sub>2</sub>O = 1:2, 2 h (O<sub>2</sub>-free condition), 90%.

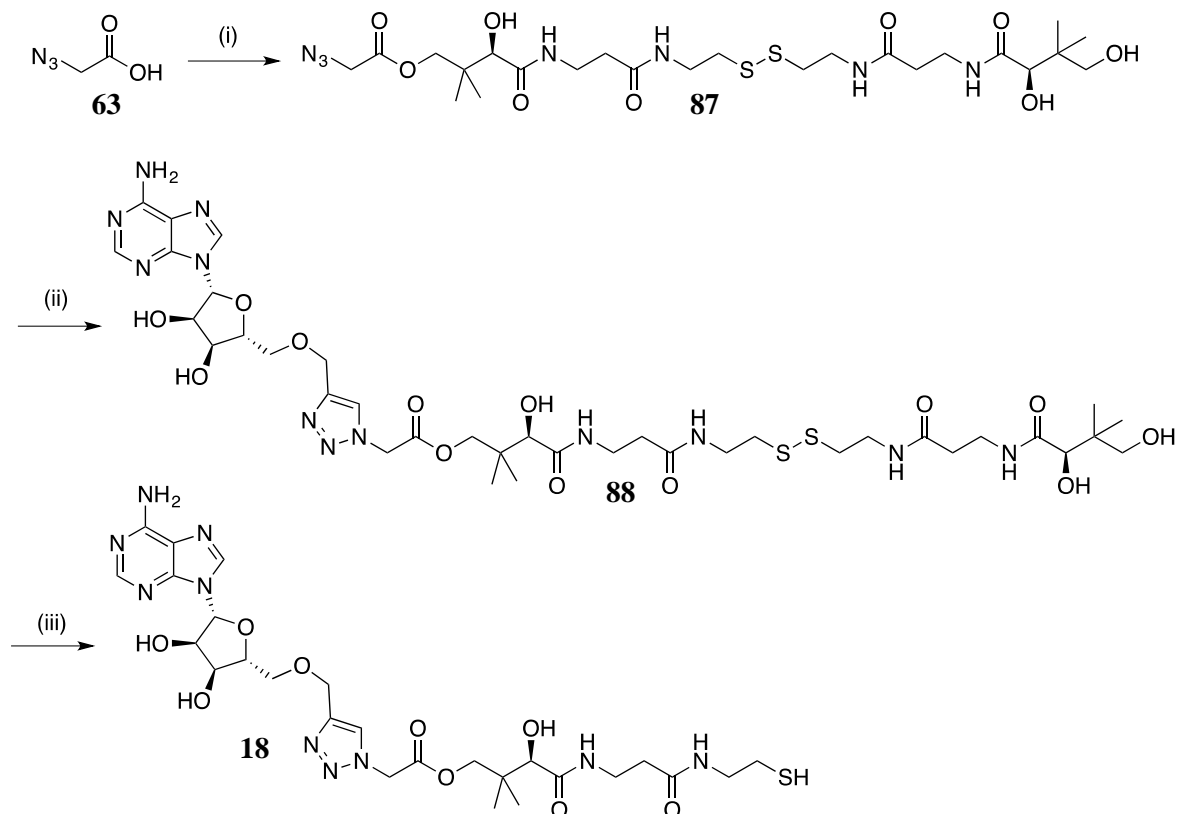
As discussed in Chapter 2, due to solubility problem of **14**, in order to increase the solubility whilst retaining the reactivity, **15** was designed. Compound **15** was prepared by CuAAC reaction between alkyne **72** and azide **86**. Preparation of **86** was achieved by amide coupling between 2-azidoacetic acid, **63** and *tert*-butyl (2-aminoethyl) carbamate, **85** in good yield, 70%. Synthesis of the carbamate, **85** was adapted from literature<sup>157, 158</sup> which reported that phenyl isocyanate reacted vigorously with aliphatic amine to provide the desired carbamate in excellent yield. Overall, synthesis of **15** starting from commercially available phenyl isocyanate afforded the product in 30% over 4 steps. (Figure 4.10)



**Figure 4.10.** Synthesis of **15**. (i) **76** (1.0 eq), CH<sub>2</sub>Cl<sub>2</sub>, 0°C to r.t., 20 h, 97%; (ii) TFA (xs), CH<sub>2</sub>Cl<sub>2</sub>, r.t., 4 h stirring then standing in fridge o/n, quantitative; (iii) **63** (0.9 eq), HOBt (1.1 eq), EDC (1.1 eq), Et<sub>3</sub>N (xs), CH<sub>2</sub>Cl<sub>2</sub>, 70%; (iv) **72** (0.9 eq), CuSO<sub>4</sub>·5H<sub>2</sub>O, THPTA, sodium ascorbate, *t*-BuOH : H<sub>2</sub>O = 1:1, 12 h (O<sub>2</sub>-free condition), 40%.

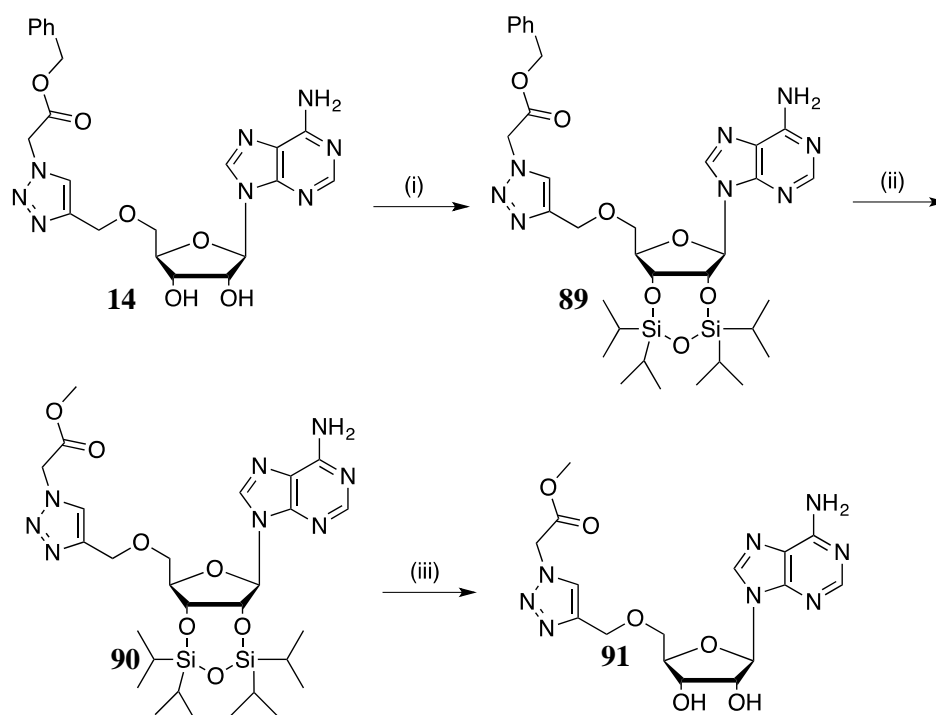
For the synthesis of an analogue with pantetheine tail, **18**, a modification of the synthesis was required. Coupling of D-pantethine and 2-azidoacetic acid using EDC and HOBt gave only the singly coupled product, **87** which was purified by HPLC. However the reaction was not in completion but the singly coupled product was purified by HPLC to bring to the next step. CuAAC reaction between azide **87** and alkyne **72** with THPTA stabilising ligand gave crude **88**

(mixed with starting material **87**). A mixture of starting material azide and product **88** was obtained by HPLC. This crude material was reduced using TCEP and purified by HPLC to obtain pure **18** (95% yield through 2 steps).



**Figure 4.11.** Synthesis of **18**. (i) D-panthethine (0.5 eq), EDC (1.2 eq), HOBT (1.2 eq), Et<sub>3</sub>N (2.0 eq), THF, CH<sub>2</sub>Cl<sub>2</sub>, 0°C to r.t, 16 h; (ii) **72** (1.0 eq), CuSO<sub>4</sub>·5H<sub>2</sub>O, THPTA, sodium ascorbate, *t*-BuOH : H<sub>2</sub>O = 1:1, 12 h (O<sub>2</sub>-free condition) (crude); (iii) TCEP (4.5 eq), H<sub>2</sub>O : Acetonitrile = 1.5 : 1, r.t, 2 h, 95% (over 2 steps).

In addition to the 5 target compounds, for the purpose of verifying the importance of  $\pi$ -stacking interaction of a phenyl ring of a ligand with Phe144, demonstrated in compound **14** and **17**, a methyl ester analogue of **14** was synthesised, **91**. For the purpose of easy handling and storage, **14** was converted to **89** by introducing protecting group TIPDS at 2' and 3' -OH of substituted adenosine ring. Literature search suggested TIPDSCl<sub>2</sub>, which would form a seven-membered ring with 2' and 3' -OH and can be taken off easily by NH<sub>4</sub>F in generally good/excellent isolated yield.<sup>155, 159</sup> The -OMe group was introduced when deprotection of -OCH<sub>2</sub>Ph group from **89** was carried out in methanol. Removal of the TIPDS protecting group of **90** afforded **91** in good yield. (Figure 4.12)



**Figure 4.12.** Synthesis of **91**. (i) TIPDSCl<sub>2</sub> (1.2 eq), pyridine, r.t, 18 h, 98%<sup>95</sup>; (ii) H<sub>2</sub>, Pd/C 10% w/w, MeOH : CH<sub>2</sub>Cl<sub>2</sub> = 1:1, r.t, 24 h, 97%; (iii) NH<sub>4</sub>F (8.6 eq), MeOH, r.t, 2 h, 86%.

### 4.3. Biological screening results of the second generation of compounds and Discussion

*In vitro* testings by Dr. Tsuchiya provided the following IC<sub>50</sub> of the six designed compounds:

**Table 4.7 – IC<sub>50</sub> testing results of the second generation of compounds**

Compound	AURKA mean IC <sub>50</sub> (μM)	S.E.M	n
<b>14</b>	7.7	0.7	3
<b>15</b>	16.7	2.3	3
<b>16</b>	27.3	5.8	3
<b>17</b>	202	n/a	1
<b>91</b>	391	n/a	1
<b>18</b>	1800	n/a	1

AURKA activity was assayed by a radiometric filter binding assay using myelin basic protein as a substrate. (See Table 1.2 legend for details). [ATP] = 5μM. n = number of experiments. n/a: not applicable. For compounds that show no inhibition or very poor inhibitory activity in preliminary assay, the exact IC<sub>50</sub> values were not pursued. S.E.M: Standard Error of Mean.

The results were in good agreement with the prediction from docking results. It was found that changing the –OPhe group in **14** to –OMe in **91** diminished the inhibition (IC<sub>50</sub> changed from 7.7 to 391 μM) This together with the computational docking results further emphasised the importance of π-stacking interaction of a phenyl ring of a ligand with Phe144. Compounds **14**

and **15** showed comparable  $IC_{50}$  to ADP, providing good indication that the acetyl triazole link can be used to replace the pyrophosphate. Comparing  $IC_{50}$  of **16** (27.3  $\mu$ M) and CoA (4.4  $\mu$ M) suggested the role of 3' phosphate/the interactions with front polar hump in AURKA inhibition. Compared the  $IC_{50}$  result of compound **16** to compound **17** which structure was the same as **16** but the –SH group was replaced by a –CH<sub>3</sub> group, it was nine times more reactive. This suggests some protein-ligand interactions involved the –SH group.

The overall conclusion would be:

- (i) Having an interaction with F144 can improve the binding of ligands in AURKA ATP binding site.
- (ii) Polar interactions with T217-E260-N261 may improve inhibitory effect and possible selectivity.
- (iii) In terms of large molecules that are able to reach the activation loop, having a polar end, or particularly –SH (to be investigated), may improve inhibition compared to a non-polar end.

Recent stability tests for compound **14** to **18** showed that only **15** was stable for 24 h. This was postulated due to the lack of an ester bond in **15**, which was present in other compounds and susceptible for hydrolysis. Therefore further development of inhibitors should avoid including ester bond in structures. The synthesis of **19** has not been carried out and the design needs a revision due to the presence of an ester bond in **19**. The retrosynthesis of **19** is presented in Appendix 4.1.

## 5. Future work

The biological data suggested that the method of designing AURKA inhibitor was in the right direction. However, in order to design new selective inhibitors with  $IC_{50}$  in the low nM range, it is essential to separate this research from CoA scaffold investigation.

The research on selectivity of CoA for AURKA should be carried out independently. Whilst this would provide valuable information on the structure of the ATP binding site and its mechanism of action, CoA itself cannot be the lead compound to design new selective inhibitors. It is proposed that the design of new selective inhibitors should start at the three regions, the hinge region, the polar region R<sub>1</sub> and the phosphate binding region R<sub>2</sub> and may grow to the pantetheine binding region R<sub>3</sub> if appropriate. (Chapter 2). Computational work suggests to further carry out synthetic chemistry in three aspects:

- (i) Phosphorylation at 3'-OH.
- (ii) Changing the head group from adenosine to another chemical moiety known in kinase inhibitors shown in Chapter 2 to increase binding at the hinge region.

- (iii) Extending the 3'-OH position to test for interaction with T217 through maleimide moiety.

On the research of the selectivity nature of CoA for AURKA, due to the lack of X-ray crystal structures and sufficient information from electron density map, more analogues of CoA should be generated in the aim to gather information about the binding mode of CoA inside the ATP binding pocket.

## 6. Synthetic experimental

All reactions were carried out under argon unless otherwise stated. Solvents and reagents were purchased from commercial suppliers and used without further purification. Tris(dibenzylideneacetone)dipalladium ( $\text{Pd}_2(\text{dba})_3$ ) was purchased from Acros Organics.

Dry solvents were purchased from Acros Organics (AcroSeal).

THPTA ligand was purchased from 2B Scientific.

Amberlite® IR120 hydrogen form was purchased from Sigma-Aldrich.

$^1\text{H}$  NMR and  $^{13}\text{C}$  NMR spectra were recorded on Bruker AMX400 (400 MHz), AMX500 (500 MHz), AMX600 (600 MHz) instruments at ambient temperature. The chemical shifts are expressed in parts per million (ppm) referenced to TMS (tetramethylsilane). Data are reported as follows:  $\delta$ , chemical shift, integration, multiplicity, coupling constants ( $J$  in Hertz, Hz), and assignment (aromatic, ar). For  $^1\text{H}$  NMR, the multiplicity used for assignment is indicated by the following abbreviations: s = singlet, d = doublet, t = triplet, q = quadruplet, quint = quintet, sext = sextet, m = multiplet, br = broad, dd = doublet of doublets and dt = doublet of triplets. For  $^{13}\text{C}$  NMR, the spectra were assigned based on the support information from  $^{13}\text{C}$ -DEPT135 and HSQC spectra.

Deuterated chloroform ( $\text{CDCl}_3$ ), dimethylsulfoxide ( $(\text{CD}_3)_2\text{SO}$ ) and methanol ( $\text{CD}_3\text{OD}$ ) were used as solvents (as stated) for all NMR analysis. For IR recordings, br= broad, m= medium, s= sharp, w= weak. A stereocentre is noted with \*.

$^{31}\text{P}$  NMR spectrum was recorded on Bruker AMX400.

Infrared spectra were obtained on a Perkin Elmer Spectrum 100 FTIR Spectrometer operating in ATR mode.

Melting points were measured with a Gallenkamp melting point apparatus.

High and low resolution mass spectrometry was performed using a VG70 SE operating in modes ES, EI or CI (+ or -) depending on the sample.

Normal phase silica gel 60 (0.04-0.063 mm, 230-400 mesh) (BDH) was used for flash column chromatography. All reactions were monitored by thin layer chromatography (TLC) unless otherwise stated. TLC plates pre-coated with silica gel 60 F-254 on aluminium (Merck KGaA) and  $\text{KMnO}_4$  as chemical stain were used, for visualisation under UV (254 or 365 nm).

*Preparative HPLC was performed on a Dionex 580 HPLC system with PDA-100 photodiode array detector, a P580 pump and a model ASI-100 automated sample injector. A Phenomenex*

*Onyx C18 100 x 10 mm column was used. Elutions were monitored at 254 nm and 280 nm and carried out according to either of the following gradients. Chromatograms were analysed using Chromeleon Software version 2.0.*

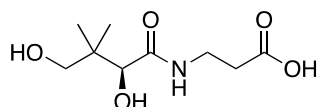
Gradient number	Composition (% of acetonitrile/water (0.1% TFA))	Duration (min)	Flow rate (mL.min <sup>-1</sup> )
1	0.5 to 1	6	2.0
2	10 to 60	20	2.0
3	7 to 37	30	2.0
4	7 to 18	18	2.0
5	10 to 98	20	2.0
6	10 to 70	15	2.0
7	15 to 40	30	2.0
8	15 to 40	30	8.0
9	15 to 20	10	8.0
10	0.1 to 0.6	8	2.0

r.t = 20°C

## 6.1. Approach A syntheses

### 6.1.1. Synthesis of the first generation of compounds

#### Pantothenic acid, **22** <sup>160</sup>



#### *From sodium salt*

The free-form pantothenic acid, **22** was obtained by dissolving sodium pantothenate, **21** (0.70 g, 2.90 mmol) in MeOH (20 mL) and passing the solution through an Amberlite IR-120 ion exchange column with distilled water until the pH reached 6.6 (approx. 1.5 L). The solvents (water and methanol) were removed *in vacuo* and the resulting colourless oil was dried under high vacuum to give **22** quantitatively.

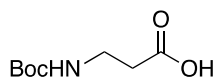
#### *From D-pantolactone and β-alanine*

To β-alanine (1.69 g, 18.78 mmol, 1.20 eq) in MeOH (90 mL) was added NH<sub>4</sub>Et<sub>2</sub> (4 mL). The solution was heated under reflux for 5 h. D-Pantolactone (1.00 g x 2, 15.21 mmol) was then added at two 12-hour intervals. The reaction solution was heated under reflux overnight. The solvent was then removed *in vacuo* and the residue was re-dissolved in water (40 mL) and passed through an Amberlite IR-120 ion exchange column with distilled water until neutrality

was achieved. The obtaining solute was concentrated *in vacuo* to about 200 mL and extracted with CH<sub>2</sub>Cl<sub>2</sub> (10 x 250 mL). The combined aqueous layer was then dried *in vacuo* to give **22** as a colourless oil (1.93 g, 58%).

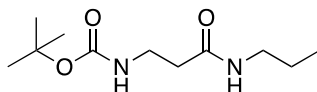
<sup>1</sup>H-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 600 MHz): 7.71 (t, 1H, *J* 6.0 Hz, NH), 5.42 (d, 1H, *J* 6.0 Hz, C\*H(OH)-CONH), 4.51 (t, 1H, *J* 6.0 Hz, OHCH<sub>2</sub>), 3.70 (d, 1H, *J* 6.0 Hz, C\*HOH), 3.15-3.33 (m, 4H, CONHCH<sub>2</sub>CH<sub>2</sub>COOH, HOCH<sub>2</sub>), 2.41 (t, 2H, *J* 7.2 Hz, CH<sub>2</sub>CH<sub>2</sub>COOH), 0.79 (s, 3H, CH<sub>3</sub>), 0.77 (s, 3H, CH<sub>3</sub>). <sup>13</sup>C-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 150 MHz): 177 (CONH), 173 (COOH), 74.5 (CH(OH)CONH), 39.6 (C(CH<sub>3</sub>)<sub>2</sub>), 34.2 (NHCH<sub>2</sub>CH<sub>2</sub>COOH, HOCH<sub>2</sub>), 33.9 (NHCH<sub>2</sub>CH<sub>2</sub>COOH), 20.3 (CH<sub>3</sub>), 19.6 (CH<sub>3</sub>). MS (CI) *m/z* (%): 221 (14), 220 (100), 202 (20). HRMS C<sub>9</sub>H<sub>17</sub>NO<sub>5</sub>: calc. 220.11850, found 220.118001, error: 2.30 ppm.

### ***N*-Boc-β-alanine, 27** <sup>161</sup>



β-Alanine (1.79 g, 20.10 mmol) was dissolved in water (20 mL) and NaOH (10 M, 10 mL) was added at 0°C. The solution was stirred at 0°C for 15 min and Boc<sub>2</sub>O (4.44 g, 20.3 mmol, 1.0 eq) was added. The reaction mixture was stirred at 15°C for 22 h. The solvent was then removed *in vacuo* to about 20 mL and then water (50 mL) was added. The mixture was extracted once with pentane (50 mL). The pH of the aqueous layer was then adjusted to 2 by HCl 9% solution. Extraction continued with EtOAc (3 x 100 mL). The combined organic layers were washed with brine (2 x 80 mL) and dried (MgSO<sub>4</sub>). The solvent was removed *in vacuo* to give **27** as a white powder (2.15 g, 57%). Mpt: 72 – 75°C. IR (*v*<sub>max</sub>, cm<sup>-1</sup>): 3440 (m, NH), 2969 (br, OH), 1738, 1702 (OC=ONH, C=OOH). <sup>1</sup>H-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 500 MHz): 6.86 (s, 1H, NH), 3.10 (q, 2H, *J* 6.9 Hz, CONH-CH<sub>2</sub>), 2.33 (t, 2H, *J* 6.9 Hz, CH<sub>2</sub>-COOH), 1.35 (s, 9H, (CH<sub>3</sub>)<sub>3</sub>C). (Identical by <sup>1</sup>H-NMR to the literature <sup>128a</sup>) <sup>13</sup>C-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 125 MHz): 172.9 (CONH), 155.5 (COOH), 77.7 (CH<sub>3</sub>)<sub>3</sub>C-O), 36.1 (CONH-CH<sub>2</sub>), 34.2 (CH<sub>2</sub>-COOH), 28.2 (CH<sub>3</sub>)<sub>3</sub>C). MS (CI) *m/z* (%): 190 (100), 134 (30). HRMS C<sub>8</sub>H<sub>15</sub>NO<sub>4</sub>: calc. 190.10793, found 190.107173, error: 4 ppm.

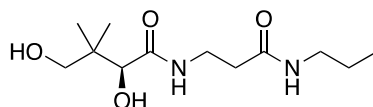
### ***tert*-Butyl-(3-oxo-3-(propylamino)propyl)carbamate, 29**



*N*-Boc-β-alanine, **27** (1.07 g, 5.65 mmol) was dissolved in THF (11 mL) at 0°C. NEt<sub>3</sub> (1.12 mL, 8.04 mmol, 1.40 eq) was added to the solution followed by iso-*tert*-butyl chloroformate (0.91 mL, 6.61 mmol, 1.20 eq). The reaction mixture was stirred from 0°C to 15°C over 3 h. Meanwhile, a solution of propylamine (0.6 mL, 7.30 mmol, 1.30 eq) and NEt<sub>3</sub> (0.35 mL, 2.51 mmol) in THF (5 mL) was stirred at 15°C over 1 h. After 3 h, the propylamine solution was

added into the reaction mixture and the whole mixture was stirred at 15°C for 17 h. The white precipitate was then filtered off and washed with EtOAc. The filtrate was reduced and underwent flash column chromatography (20 to 25% EtOAc in Hexane) to give **29** as a white powder (0.75 g, 58%). R<sub>f</sub>: 0.53 (EtOAc : Hex = 1:1). Mpt: 59 – 60°C. IR (ν<sub>max</sub>, cm<sup>-1</sup>): 3324 (m, NH), 2930 (w, sp<sup>3</sup>-CH), 1688, 1641 (OC=ONH, C=ONH). <sup>1</sup>H-NMR (CDCl<sub>3</sub>, 500 MHz): 5.74 (br, 1H, NH), 5.13 (br, 1H, NH), 3.40 (t, 2H, *J* 6.0 Hz, C(*t*-Bu)OCONH-CH<sub>2</sub>), 3.21 (q, 2H, *J* 7.1 Hz, CONH-CH<sub>2</sub>), 2.40 (t, 2H, *J* 6.0 Hz, C(*t*-Bu)OCONH-CH<sub>2</sub>-CH<sub>2</sub>-CONH), 1.49-1.56 (m, 2H, CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>), 1.43 (s, 9H, (CH<sub>3</sub>)<sub>3</sub>C), 0.94 (t, 3H, *J* 7.3 Hz, CH<sub>2</sub>CH<sub>3</sub>). <sup>13</sup>C-NMR (CDCl<sub>3</sub>, 125 MHz): 171.7 (CONH), 156.5 (COOH), 79.7 (CH<sub>3</sub>)<sub>3</sub>C-O), 41.5 (CONH-CH<sub>2</sub>), 36.8 (C(*t*-Bu)OCONH-CH<sub>2</sub>), 36.5 (C(*t*-Bu)OCONH-CH<sub>2</sub>-CH<sub>2</sub>-CONH), 28.2 (CH<sub>3</sub>)<sub>3</sub>C), 22.9 (CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>), 11.5 (CH<sub>2</sub>CH<sub>3</sub>). MS (CI) *m/z* (%): 231 (21), 203 (18), 175 (54), 174 (32), 157 (18), 131 (100). HRMS C<sub>11</sub>H<sub>22</sub>N<sub>2</sub>O<sub>3</sub>: calc. 231.17087, found 231.17120, error: 1.43 ppm.

**(*R*)-2,4-dihydroxy-3,3-dimethyl-*N*-(3-oxo-3(propylamino)propyl)butanamide, **20**** <sup>120</sup>



*Route 1*

The free-form pantothenic acid, **22** (1.93 g, 8.81 mmol) was dissolved in dry DMF (10 mL) and the solution was cooled to 0°C. DPPA (3.0 mL, 13.50 mmol, 1.50 eq) was added to the solution, followed by NEt<sub>3</sub> (2.5 mL, 18.10 mmol, 2.10 eq). The solution was stirred at 0°C for 45 min then propylamine (1.15 mL, 13.71 mmol, 1.60 eq) was added at 0°C. The reaction mixture was continued stirring at 0°C for 3 h then at r.t for 17 h. The solvent was then removed *in vacuo* to give a viscous orange yellow oil. Purification by flash column chromatography (0 to 2% MeOH in EtOAc) gave the product **20** as a colourless oil (1.02 g, 47%).

*Route 2 - 2 steps*

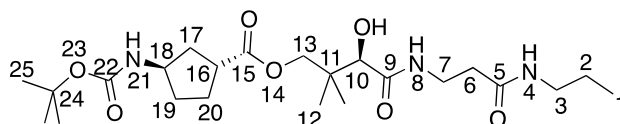
*tert*-Butyl (3-oxo-3-(propylamino)propyl)carbamate, **29** (0.29 g, 1.26 mmol) was deprotected by stirring with TFA (2.5 mL) in CH<sub>2</sub>Cl<sub>2</sub> (10 mL) for 6 h at 15°C. The solvent was removed *in vacuo* to give a colourless oil that was dried *in vacuo* for 2 days. The deprotected product, 3-amino-*N*-propylpropanamide was confirmed by NMR analysis and the yield was quantitative. It was then taken to the next step without further purification.

To 3-amino-*N*-propylpropanamide (0.16 g, 1.26 mmol) in MeOH (13 mL) NHEt<sub>2</sub> (2.0 mL, 19 mmol, xs) was added. The solution was heated under reflux for 7 h then D-pantolactone, **26** (0.16 g, 1.25 mmol, 0.99 eq) was added. The mixture was heated under reflux for 22 h. Purification by flash column chromatography (0 to 1.7% MeOH in EtOAc) yielded **20** as a colourless oil (0.24 g, 92 %).



R<sub>f</sub>: 0.26 (10% MeOH in EtOAc). <sup>1</sup>H-NMR (CDCl<sub>3</sub>, 600 MHz): 7.54 (br, 1H, NH), 6.42 (br, 1H, NH), 4.00 (s, 1H, CHOH), 3.54 (br and m, 2H, CH(OH)CONHCH<sub>2</sub>), 3.48-3.49 (m, br, 2H, OHCH<sub>2</sub>), 3.16 (q, 2H, *J* 6.7 Hz, CONHCH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>), 2.42 (t, 2H, *J* 6.0 Hz, CH(OH)CONHCH<sub>2</sub>CH<sub>2</sub>), 1.51 (sext, 2H, *J* 7.2 Hz, CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>), 0.98 (s, 3H, C(CH<sub>3</sub>)<sub>2</sub>), 0.95 (s, 3H, C(CH<sub>3</sub>)<sub>2</sub>), 0.88 (t, 3H, *J* 7.2 Hz, CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>). (Identical by <sup>1</sup>H-NMR to the literature <sup>120</sup>) <sup>13</sup>C-NMR (CDCl<sub>3</sub>, 150 MHz): 173.4 (CONH), 171.4 (COOH), 77.9 (CH(OH)CONH), 71.0 (OHCH<sub>2</sub>), (OHCH<sub>2</sub>), 51.0 (CH(OH)CONHCH<sub>2</sub>CH<sub>2</sub>CONH), 41.5 (CONHCH<sub>2</sub>), 39.4 (C(CH<sub>3</sub>)<sub>2</sub>), 35.8 (CH(OH)CONHCH<sub>2</sub>CH<sub>2</sub>CONH), 22.9 (CH<sub>2</sub>CH<sub>3</sub>), 21.6 (C(CH<sub>3</sub>)<sub>2</sub>), 20.6 (C(CH<sub>3</sub>)<sub>2</sub>), 11.5 (CH<sub>2</sub>CH<sub>3</sub>). MS (CI) *m/z* (%): 259 (100). HRMS C<sub>12</sub>H<sub>24</sub>N<sub>2</sub>O<sub>4</sub>: calc. 259.1658, found 259.1658, error: 0.00 ppm.

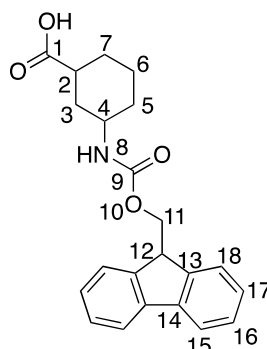
**(*R*)-3-hydroxy-2,2-dimethyl-4-oxo-4-((3-oxo-3-(propylamino)propyl)amino)butyl (*1R,3R*)-3-((*tert*-butoxycarbonyl)amino)cyclopentane-1-carboxylate, **31****



To a solution of commercially available (*1R, 3R*)-*N*-BOC 1-aminocyclopentane 3-carboxylic acid (0.42 g, 1.74 mmol, 1.1 eq) in THF (5 mL) at 0°C, was added DCC (0.66 g 3.17 mmol, 2.0 eq). The mixture was stirred at 0°C and then allowed to gradually warm up to r.t for 2 h before a catalytic amount of DMAP (77 mg) was added. The mixture was then stirred at r.t for 30 min. (*R*)-2,4-dihydroxy-3,3-dimethyl-*N*-(3-oxo-3(propylamino)propyl)butanamide, **20** (0.41 g, 1.56 mmol) in THF (7 mL) was added to the reaction mixture and the solution was stirred at 45°C for 15 h. The precipitate was filtered off and the filtrate was concentrated *in vacuo*. Purification by flash column chromatography (EtOAc : Hexane = 1:1, then EtOAc : Hexane = 3:1, 100% EtOAc then 0.5% MeOH in EtOAc) gave the product, **31** as colourless flakes upon drying *in vacuo* (0.25 g, 33 %). R<sub>f</sub>: 0.53 (10% MeOH in EtOAc). Mpt: 38 – 40°C. IR (ν<sub>max</sub>, cm<sup>-1</sup>): 3310 (m, NH), 2966 (m, sp<sup>3</sup>-CH), 1688, 1645 (OC=ONH, C=ONH). <sup>1</sup>H-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 600 MHz): 3.90-4.03 (br, m, 1H, H-18), 3.87 (s, 1H, CH, H-10), 3.41-3.51 (m, 2H, CH<sub>2</sub>, H-7), 4.05 (d, 1H, CH<sub>2</sub>, *J* 10.65 Hz, H-13), 3.97 (d, 1H, CH<sub>2</sub>, *J* 10.65 Hz, H-13), 3.12 (t, 2H, CH<sub>2</sub>, *J* 7.0 Hz, H-3), 2.96 (quint, 1H, CH, *J* 8.0 Hz, H-16), 2.42 (t, 2H, *J* 6.8 Hz, CH<sub>2</sub>, H-6), 2.15-2.29 (m, 1H, CH<sub>2</sub>, H-19), 2.09-2.18 (m, 2H, CH<sub>2</sub>, H-17 and H-19), 1.74-1.84 (m, 2H, CH<sub>2</sub>, H-17 and H-20), 1.48-1.53 (m, 3H, CH<sub>2</sub>, H-20 and H-2), 1.43 (s, 9H, 3 x CH<sub>3</sub>, H-25), 0.97 (s, 3H, CH<sub>3</sub>, H-12), 0.95 (s, 3H, CH<sub>3</sub>, H-12), 0.90 (t, 3H, CH<sub>3</sub>, *J* 7.2 Hz, H-1). <sup>13</sup>C-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 150 MHz): 177.5 (C=O, C-22), 176.7 (C=O, C-15), 175.5 (C=O, C-9), 173.6 (C=O, C-5), 79.9 (C, C-24), 75.9 (CH, C-10), 71.3 (CH<sub>2</sub>, C-13), 53.1 (CH, C-18), 43.3 (CH, C-16), 42.2 (CH<sub>2</sub>, C-3), 39.5 (C, C-11), 37.2 (CH<sub>2</sub>, C-7), 36.8 (CH<sub>2</sub>, C-19), 33.8 (CH<sub>2</sub>, C-20), 29.2 (CH<sub>2</sub>, C-17), 23.6 (CH<sub>2</sub>, C-2), 21.7 (CH<sub>3</sub>, C-12), 21.6 (CH<sub>3</sub>, C-12), 21.0 (CH<sub>3</sub>, C-25), 11.8 ((CH<sub>3</sub>, C-1). MS (CI) *m/z* (%): 472 (15), 416 (12), 398 (12), 370 (44), 258 (52), 243 (63), 214 (36), 188 (46), 157

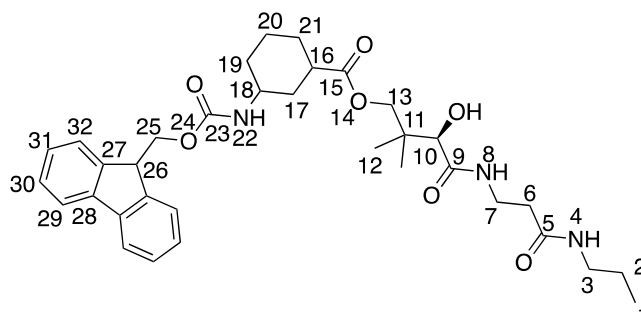
(58), 131 (97), 122 (100). HRMS  $C_{23}H_{42}N_3O_7$ : calc 472.30173, found 472.300985, error: 1.60 ppm.

**3-(((9*H*-fluoren-9-yl)methoxy)carbonyl)amino)cyclohexane-1-carboxylic acid, 34** <sup>162</sup>



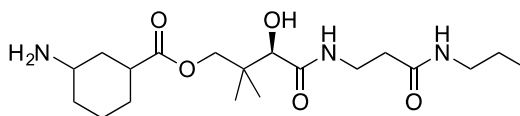
To a mixture of commercially available 3-aminocyclohexane-1-carboxylic acid (*cis*- and *trans*-isomers with >95% *cis*-3-aminocyclohexane-1-carboxylic acid) (853 mg, 5.66 mmol) and  $NaHCO_3$  (820 mg, 9.76 mmol, xs) in water (20mL) at r.t, was added FmocOSu (2.38 g, 6.77 mmol, 1.20 eq) in THF (35 mL). The reaction mixture is stirred at r.t for 30 h. THF was then removed *in vacuo* and a white precipitate started to crush out. EtOAc (150 mL) was added to the mixture then HCl (3 M, 70 mL) was added slowly to the solution whilst stirring. The white precipitate was dissolved into organic layer to form a milky organic layer, which was extracted out. The organic layer was washed twice with brine (2 x 200 mL). To the milky organic layer, was added THF (100 mL) to form a clear solution, which was then dried ( $MgSO_4$ ). The solvent was removed *in vacuo* to give the product as a white powder (2.06 g, 95%). R<sub>f</sub>: 0.71 (10% MeOH in EtOAc). IR ( $\nu_{max}$ ,  $cm^{-1}$ ): 3323 (m, NH), 3067 (br, OH), 2939, 2860(m,  $sp^3$ -CH), 1679 (s, OC=ONH and C=OOH), 1537 (s, OCONH). <sup>1</sup>H-NMR ( $CD_3OD$ , 600 MHz): 7.80 (d, 2H, Ar-H, *J* 7.80 Hz, H-15), 7.60 (d, 2H, Ar-H, *J* 7.80 Hz, H-18), 7.40 (t, 2H, Ar-H, *J* 7.20 Hz, H-16), 7.31 (t, 2H, Ar-H, *J* 7.20 Hz, H-17), 4.36 (d, 2H, CH<sub>2</sub>, *J* 6.11 Hz, H-11), 4.21 (t, 1H, CH, *J* 7.00 Hz, H-12), 3.40-3.42 (br, m, 1H, CH, H-4), 2.34-2.38 (br, m, 1H, CH, H-2), 2.12-2.16 (br, m, 1H, CH<sub>2</sub>, H-3), 1.83-1.92 (br, m, 2H, CH<sub>2</sub>, H-7), 1.12-1.40 (m, 5H, CH<sub>2</sub>, H-3 and H-5 and H-6). <sup>13</sup>C-NMR ( $CD_3OD$ , 150 MHz): 176.3 (C=O, C-1), 155.3 (C=O, C-9), 143 (Ar-C, C-13), 139 (Ar-C, C-14), 128.8 (Ar-C, C-16), 128 (Ar-C, C-15), 126 (Ar-C, C-17), 121 (Ar-C, C-18), 65.4 (CH<sub>2</sub>, C-11), 49.0 (CH, C-4), 46.8 (CH, C-12), 42.0 (CH, C-2), 35.4 (CH<sub>2</sub>, C-3), 32.1, 28.1 and 25.2 (CH<sub>2</sub>, C-5, C-6 and C-7). HRMS  $C_{22}H_{24}NO_4$ : calc 366.1705, found 366.1718, error: 3.60 ppm.

**(R)-3-hydroxy-2,2-dimethyl-4-oxo-4-((3-oxo-3-(propylamino)propyl)amino)butyl 3-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)cyclohexane-1-carboxylate, 36**



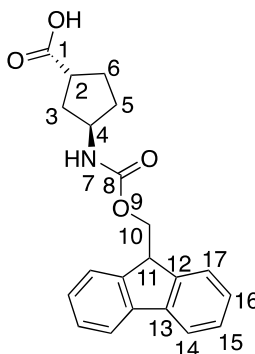
To 3-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)cyclohexane-1-carboxylic acid, **34** (390 mg, 1.07 mmol) in THF (10 mL) at 0°C, DCC (590 mg, 2.83 mmol, 2.64 eq) and DMAP (80.0 mg, 0.65 mmol, 0.60 eq) were added. The mixture was stirred at 0°C for 1 h before (R)-2,4-dihydroxy-3,3-dimethyl-N-(3-oxo-3-(propylamino)propyl)butanamide, **20** (375 mg, 1.44 mmol, 1.35 eq) in THF (8 mL) was added. The solution was then stirred at 45°C for 18 h before the solvent was removed *in vacuo*. Purification by flash column chromatography (Petroleum Ether : EtOAc = 3:1 to 1:1 then 0.5% MeOH in EtOAc) yielded the product as a light yellow solid (287 mg, 37%). R<sub>f</sub>: 0.50 (10% MeOH in EtOAc). <sup>1</sup>H-NMR (CD<sub>3</sub>OD, 600 MHz): 7.80 (d, 2H, Ar-H, *J* 7.80 Hz, H-29), 7.60 (d, 2H, Ar-H, *J* 7.80 Hz, H-32), 7.40 (t, 2H, Ar-H, *J* 7.20 Hz, H-30), 7.31 (t, 2H, Ar-H, *J* 7.20 Hz, H-31), 4.34 (d, 2H, CH<sub>2</sub>, *J* 7.20 Hz, H-25), 4.20 (t, 1H, CH, *J* 7.20 Hz, H-26), 4.05 (d, 1H, CH<sub>2</sub>, *J* 10.62 Hz, H-13), 3.92 (d, 1H, CH<sub>2</sub>, *J* 10.62 Hz, H-13), 3.87 (s, 1H, CH, H-10), 3.41-3.50 (m, 3H, CH and CH<sub>2</sub>, H-18 and H-7), 3.11 (t, 2H, CH<sub>2</sub>, *J* 7.20 Hz, H-3), 2.39-2.42 (m, br, 3H, CH and CH<sub>2</sub>, H-16 and H-6), 2.12-2.15 (br, 1H, CH<sub>2</sub>, H-17), 1.83-1.89 (br, 2H, CH<sub>2</sub>, H-21), 1.50-1.51 (m, 2H, CH<sub>2</sub>, H-2), 1.12-1.40 (m, 5H, CH<sub>2</sub>, H-17 and H-19 and H-20), 0.98 (s, 3H, CH<sub>3</sub>, H-12), 0.94 (s, 3H, CH<sub>3</sub>, H-12), 0.89 (t, 3H, CH<sub>3</sub>, *J* 7.20 Hz, H-1). <sup>13</sup>C-NMR (CD<sub>3</sub>OD, 150 MHz): 176.5 (C=O, C-15), 175.4 (C=O, C-9), 173.6 (C=O, C-5), 158.0 (C=O, C-23), 145.3 (Ar-C, C-27), 142.6 (Ar-C, C-28), 128.8 (Ar-C, C-30), 128 (Ar-C, C-29), 126 (Ar-C, C-31), 121 (Ar-C, C-32), 75.9 (CH, C-10), 71.2 (CH<sub>2</sub>, C-13), 67.5 (CH<sub>2</sub>, C-25), 49.1 (CH, C-18), 48.8 (CH, C-26), 43.8 (CH, C-16), 42.2 (CH<sub>2</sub>, C-3), 39.6 (C, C-11), 36.8 (CH<sub>2</sub>, C-7), 36.4 (CH<sub>2</sub>, C-6), 36.3 (CH<sub>2</sub>, C-17), 29 (CH<sub>2</sub>, C-19 and C-20), 27 (CH<sub>2</sub>, C-21), 23.6 (CH<sub>2</sub>, C-2), 20.3 (CH<sub>3</sub>, C-12), 19.8 (CH<sub>3</sub>, C-12), 11.8 (CH<sub>3</sub>, C-1). HRMS C<sub>34</sub>H<sub>46</sub>N<sub>3</sub>O<sub>7</sub>: calc 608.3336, found 608.3332, error: 0.66 ppm.

**(R)-3-hydroxy-2,2-dimethyl-4-oxo-4-((3-oxo-3-(propylamino)propyl)amino)butyl 3-aminocyclohexane-1-carboxylate, 37**



(R)-3-hydroxy-2,2-dimethyl-4-oxo-4-((3-oxo-3-(propylamino)propyl)amino)butyl 3-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)cyclohexane-1-carboxylate, **36** (30 mg, 0.05 mmol) was dissolved in THF (2.5 mL) at r.t and piperidine (1.0 mL, 0.01 mol, xs) was added. The solution was stirred at r.t for 4 h before THF and piperidine were removed *in vacuo*. The mixture triturated until no trace of UV active spot was observed in toluene. The light yellow oil left after toluene wash was the product, which was dried *in vacuo* (1.5 h) (12.0 mg, 67%) and was used immediately in the next step without further purification.  $R_f$ : 0 (30% MeOH in EtOAc + ammonia).

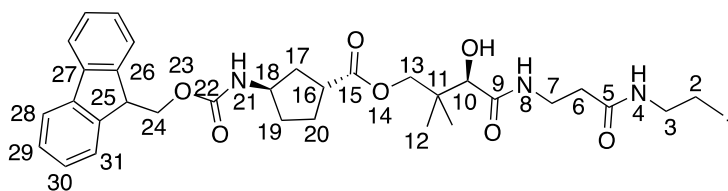
**(1R,3R)-3-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)cyclopentane-1-carboxylic acid, 35**<sup>163</sup>



(1R,3R)-3-((*tert*-butoxycarbonyl)amino)cyclopentane-1-carboxylic acid (269 mg, 1.11 mmol) was suspended in  $\text{CH}_2\text{Cl}_2$  (8.0 mL) and TFA (4.0 mL). The mixture was stirred at r.t for 15 h then dried *in vacuo* for 4 h before re-dissolving in THF (12 mL). To this mixture, were added a solution of  $\text{NaHCO}_3$  (163 mg, 1.94 mmol, 1.75 eq) in  $\text{H}_2\text{O}$  (7 mL) and FmocOSu (489 mg, 1.39 mmol, 1.25 eq). The reaction mixture was stirred at r.t for 22 h then THF was removed *in vacuo*. EtOAc (25 mL) was added along with HCl (3 M, 25 mL). The mixture was stirred at r.t until all the white precipitate was dissolved, forming two layers. After extraction with EtOAc (3 x 25 mL), the combined organic layers were washed with brine once. It was then dried ( $\text{MgSO}_4$ ). The solvent was removed *in vacuo* to provide the crude product as a light yellow oil, which solidified gradually under high vacuum. The crude oil was then washed several time with  $\text{Et}_2\text{O}$  and dried *in vacuo* to give the product as a white solid (391 mg, quantitative).  $R_f$ : 0.24 (Petroleum Ether : EtOAc = 1:1). IR ( $\nu_{\text{max}}$ ,  $\text{cm}^{-1}$ ): 3317, 3063 (br, -OH), 2956 (br,  $\text{sp}^3\text{-CH}$ ), 1739, 1685 (s, C=O, OC=ONH).  $^1\text{H-NMR}$  ( $\text{CD}_3\text{OD}$ , 600 MHz): 7.83 (d, 2H, Ar-H,  $J$  7.80 Hz, H-14), 7.64 (d, 2H, Ar-H,  $J$  7.80 Hz, H-17), 7.42 (t, 2H, Ar-H,  $J$  7.20 Hz, H-15), 7.32 (t, 2H, Ar-H,  $J$  7.20 Hz, H-16), 4.36 (d, 2H,  $\text{CH}_2$ ,  $J$  6.60 Hz, H-10), 4.19 (t, 1H, CH,  $J$  6.60 Hz, H-11),

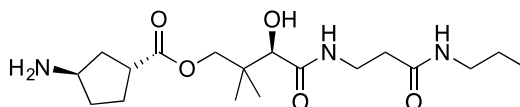
4.00-4.01 (br, m, 1H, **CH**, H-4), 2.88-2.92 (m, 1H, **CH**, H-2), 2.12-2.13 (m, 1H, **CH<sub>2</sub>**, H-5), 1.99-2.09 (br, m, 2H, **CH<sub>2</sub>**, H-3 and H-6), 1.76-1.83 (br, m, 2H, **CH<sub>2</sub>**, H-3 and H-5), 1.51-1.55 (m, 1H, **CH<sub>2</sub>**, H-6). <sup>13</sup>C-NMR (CD<sub>3</sub>OD, 150 MHz): 179 (**C=O**, C-1), 158 (**C=O**, C-8), 145.4 (Ar-**C**, C-12), 142.6 (Ar-**C**, C-13), 129 (Ar-**C**, C-15), 128 (Ar-**C**, C-14), 126 (Ar-**C**, C-16), 121 (Ar-**C**, C-17), 67.4 (**CH<sub>2</sub>**, C-10), 53.5 (**CH**, C-4), 49 (**CH**, C-11), 43.0 (**CH**, C-2), 37.1 (**CH<sub>2</sub>**, C-5), 33.4 (**CH<sub>2</sub>**, C-6), 28.9 (**CH<sub>2</sub>**, C-3). (Identical by <sup>1</sup>H- and <sup>13</sup>C-NMR to the literature <sup>163</sup>) HRMS C<sub>21</sub>H<sub>22</sub>NO<sub>4</sub>: calc 352.1549, found 352.1545, error: 1.1 ppm.

**(R)-3-hydroxy-2,2-dimethyl-4-oxo-4-(((3-oxo-3-(propylamino)propyl)amino)butyl (1R,3R)-3-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)cyclopentane-1-carboxylate, 38**



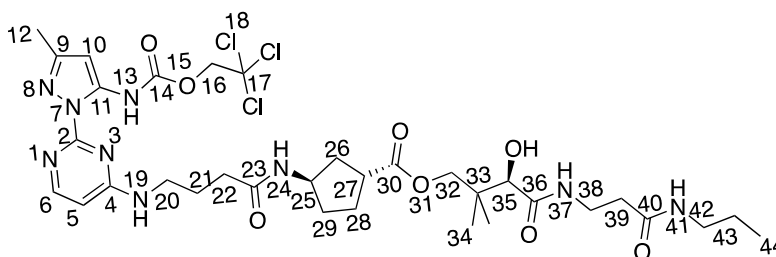
To a solution of (1R,3R)-3-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)cyclopentane-1-carboxylic acid, **35** (268 mg, 0.76 mmol) in THF (9.0 mL) at 0°C, DCC (349 mg, 1.69 mmol, 2.22 eq) and DMAP (20 mg, 0.16 mmol) were added. The reaction mixture was stirred at 0°C for 1 h. Then **20** (243 mg, 0.93 mmol, 1.22 eq) in THF (6.0 mL) was added and the reaction mixture was stirred at 45°C for 24 h. The precipitate was filtered. The filtrate was purified by flash column chromatography (Petroleum Ether : EtOAc = 3:1 then 100% EtOAc then 0.4 to 0.8% MeOH in EtOAc) to give **38** as a white solid (145 mg, 32%). R<sub>f</sub>: 0.61 (10% MeOH in EtOAc). <sup>1</sup>H-NMR (CD<sub>3</sub>OD, 600 MHz): 7.80 (d, 2H, Ar-**H**, *J* 7.80 Hz, H-28), 7.64 (d, 2H, Ar-**H**, *J* 7.80 Hz, H-31), 7.40 (t, 2H, Ar-**H**, *J* 7.20 Hz, H-29), 7.30 (t, 2H, Ar-**H**, *J* 7.20 Hz, H-30), 4.36 (d, 2H, **CH<sub>2</sub>**, *J* 6.60 Hz, H-24), 4.20 (t, 2H, *J* 6.60 Hz, **CH**, H-25), 4.00-4.25 (m, 2H, **CH** and **CH<sub>2</sub>**, H-13 and H-18), 3.83-3.97 (m, 1H, 2H, **CH** and **CH<sub>2</sub>**, H-13 and H-10), 3.43-3.57 (m, 2H, **CH<sub>2</sub>**, H-7), 3.10-3.13 (m, 2H, **CH<sub>2</sub>**, H-3), 3.00 (br, m, 1H, **CH**, H-16), 2.40-2.43 (m, 2H, **CH<sub>2</sub>**, H-6), 2.15-2.29 (m, 1H, **CH<sub>2</sub>**, H-19), 1.85-2.15 (m, 2H, **CH<sub>2</sub>**, H-17 and H-20), 1.65-1.85 (m, 2H, **CH<sub>2</sub>**, H-17 and H-19), 1.48-1.53 (m, 3H, **CH<sub>2</sub>**, H-20 and H-2), 0.97 (s, 3H, **CH<sub>3</sub>**, H-12), 0.95 (s, 3H, **CH<sub>3</sub>**, H-12), 0.90 (t, 3H, **CH<sub>3</sub>**, *J* 7.20 Hz, H-1). <sup>13</sup>C-NMR (CD<sub>3</sub>OD, 150 MHz): 176.7 (**C=O**, C-15), 175.4 (**C=O**, C-9), 173.6 (**C=O**, C-5), 158.3 (**C=O**, C-22), 145.4 (Ar-**C**, C-26), 142.6 (Ar-**C**, C-27), 128.8 (Ar-**C**, C-29), 128 (Ar-**C**, C-28), 126 (Ar-**C**, C-30), 121 (Ar-**C**, C-31), 75.9 (**CH**, C-10), 71.3 (**CH<sub>2</sub>**, C-13), 69 (**CH<sub>2</sub>**, C-24), 53.5 (**CH**, C-18), 49 (**CH**, C-25), 43.8 (**CH**, C-16), 42.2 (**CH<sub>2</sub>**, C-3), 39.6 (**C**, C-11), 37.6 (**CH<sub>2</sub>**, C-6), 37.4 (**CH<sub>2</sub>**, C-7), 37.1 (**CH<sub>2</sub>**, C-19), 33.8 (**CH<sub>2</sub>**, C-20), 29.4 (**CH<sub>2</sub>**, C-17), 23.5 (**CH<sub>2</sub>**, C-2), 21.7 (**CH<sub>3</sub>**, C-12), 21.5 (**CH<sub>3</sub>**, C-12), 11.7 (**CH<sub>3</sub>**, C-1). HRMS C<sub>33</sub>H<sub>43</sub>N<sub>3</sub>O<sub>7</sub>: calc 594.3136, found 594.3179, error: 7.2 ppm.

**(R)-3-hydroxy-2,2-dimethyl-4-oxo-4-((3-oxo-3-(propylamino)propyl)amino)butyl (1R,3R)-3-aminocyclopentane-1-carboxylate, 32**



To a solution of **38** (276 mg, 0.46 mmol) in THF (15 mL) at r.t, was added piperidine (7.0 mL, xs). The reaction mixture was stirred at r.t for 7 h before THF and piperidine were removed *in vacuo*. The mixture was triturated with toluene until no trace of UV active spot was observed in toluene. The colourless oil left after toluene wash was the product, which was dried *in vacuo* (12 h) (138 mg, 80%) and was used straight to the next step without further purification.

**(R)-3-hydroxy-2,2-dimethyl-4-oxo-4-((3-oxo-3-(propylamino)propyl)amino)butyl (1R,3R)-3-(4-((2-(3-methyl-5-(((2,2,2-trichloroethoxy)carbonyl)amino)-1H-pyrazol-1-yl)pyrimidin-4-yl)amino)butanamido)cyclopentane-1-carboxylate, 40**

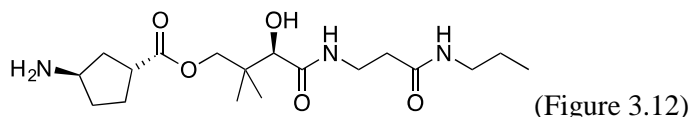


4-((2-(3-methyl-5-(((2,2,2-trichloroethoxy)carbonyl)amino)-1H-pyrazol-1-yl)pyrimidin-4-yl)amino)butanoic acid, **39** (60mg, 0.13 mmol) was dissolved in DMF (2.2 mL) at 0°C. HOAt (33 mg, 0.24 mmol, 1.86 eq) and PyAOP (245 mg, 0.47 mmol, 3.60 eq) and DIPEA (0.12 mL, xs) were added to the reaction mixture followed by **32** (64 mg, 0.17 mmol, 1.30 eq) in DMF (2.5 mL). The reaction mixture was stirred from 0°C to r.t gradually and kept stirring at r.t for 48 h. Purification by flash column chromatography (4 to 20% MeOH in CH<sub>2</sub>Cl<sub>2</sub>) yielded crude mixture (44 mg) (checked by LCMS: MW = 804 at 2.12 min). The crude mixture was purified by preparative HPLC *Gradient 10*, product peak at 3.80 min to afford the product as a white solid (18 mg, 17%). R<sub>f</sub>: 0.46 (15% MeOH in CH<sub>2</sub>Cl<sub>2</sub>+DIPEA). <sup>1</sup>H-NMR (CD<sub>3</sub>OD, 600 MHz): 8.01 (d, 1H, Ar-H, *J* 6.00 Hz, H-6), 6.50 (s, Ar-H, H-10), 6.38 (d, Ar-H, *J* 6.00 Hz, H-5), 5.00 (s, 2H, CH<sub>2</sub>, H-16), 4.18-4.20 (br, 1H, CH, H-25), 4.05 (d, 1H, CH<sub>2</sub>, *J* 6.00 Hz, H-32), 3.83-3.97 (m, 2H, CH and CH<sub>2</sub>, H-32 and H-35), 3.42-3.53 (m, 4H, CH<sub>2</sub>, H-38 and H-20), 3.11 (t, 2H, CH<sub>2</sub>, *J* 6.00 Hz, H-42), 2.92-2.94 (br, 1H, CH, H-27), 2.41 (t, 2H, CH<sub>2</sub>, *J* 6.00 Hz, H-39), 2.28-2.30 (m, 5H, CH<sub>3</sub> and CH<sub>2</sub>, H-12 and H-22), 2.13-2.15 (m, 1H, CH<sub>2</sub>, H-19), 1.95-2.02 (m, 4H, CH<sub>2</sub>, H-21 and H-26 (1) and H-28 (1)), 1.72-1.82 (m, 2H, CH<sub>2</sub>, H-26 and H-29), 1.43-1.53 (m, 3H, CH<sub>2</sub>, H-28 and H-43), 0.97 (s, 3H, CH<sub>3</sub>, H-34), 0.95 (s, 3H, CH<sub>3</sub>, H-34), 0.91 (t, 3H, CH<sub>3</sub>, *J* 7.20 Hz, H-44). <sup>13</sup>C-NMR (CD<sub>3</sub>OD, 150 MHz): 177.3 (C=O, C-30), 175.4 (C=O, C-36), 174.9 (C=O, C-23), 173.6 (C=O, C-41), 164.1 (Ar-C, C-4), 158.2 (Ar-C, C-2), 155.2 ((Ar-C, C-6), 152.8 (Ar-C, C-9), 151.4 (C=O, C-14), 141.8 (Ar-C, C-11), 104.6 (Ar-C, C-5), 97.2 (Ar-

C, C-10), 96.5 (C, C-17), 75.8 (CH<sub>2</sub>, C-16 and C-35), 71.3 (CH<sub>2</sub>, C-32), 51.9 (CH, C-25), 43.4 (CH, C-27), 42.2 (CH<sub>2</sub>, C-42), 39.5 (C, C-33), 36.9 (CH<sub>2</sub>, C-29), 36.5 (CH<sub>2</sub>, C-38 and C-39), 34.3 (CH<sub>2</sub>, C-22), 33.2 (CH<sub>2</sub>, C-28), 26.1 (CH<sub>2</sub>, C-26), 23.6 (CH<sub>2</sub>, C-43), 21.6 (CH<sub>3</sub>, C-21), 20.8 (CH<sub>3</sub>, C-34), 20.6 (CH<sub>3</sub>, C-34), 11.4 (CH<sub>3</sub>, C-44). MS (ESI+) *m/z* (%): 806 (100), 549 (35) MS (ESI+) C<sub>33</sub>H<sub>49</sub>Cl<sub>3</sub>N<sub>9</sub>O<sub>8</sub><sup>+</sup>: calc 806.2735, found 806.2883, error: 18 ppm.

### Failed synthetic attempts

#### **(R)-3-hydroxy-2,2-dimethyl-4-oxo-4-((3-oxo-3-(propylamino)propyl)amino)butyl (1R,3R)-3-aminocyclopentane-1-carboxylate, 32**



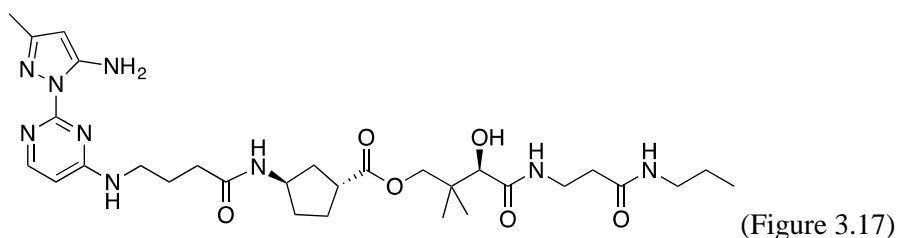
#### *Procedure 1*

To a solution of **31** (60 mg, 0.18 mmol) in DMF (3.0 mL) at r.t, was added TFA (0.3 mL, 3.90 mmol, xs). The reaction mixture was stirred at r.t for 15 h. No product was detected and starting material **31** degraded.

#### *Procedure 2*

Commercially available Amberlyst-15 was cleaned by soaking in MeOH (24 h) then in a solution of NH<sub>3</sub> 4 M in MeOH (1 h) followed by soaking in a solution of HCl 3 M in MeOH (2 h). The clean acidified Amberlyst-15 (35 mg) was added to a solution of **31** (30 mg, 0.06 mmol) in THF (2.0 mL) at r.t. The reaction mixture was stirred at r.t for 15 h. TLC (10 to 15% MeOH in EtOAc) showed the degrading starting material but no clear signal of product.

#### **(R)-3-hydroxy-2,2-dimethyl-4-oxo-4-((3-oxo-3-(propylamino)propyl)amino)butyl (1R,3R)-3-(4-((2-(5-amino-3-methyl-1H-pyrazol-1-yl)pyrimidin-4-yl)amino)butanamido)cyclopentane-1-carboxylate, 13**

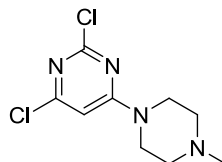


To a solution of **40** (18 mg, 2.20 x 10<sup>-2</sup> mmol) in THF (4 mL) added NH<sub>4</sub>OAc (0.25 mL) and Zn dust (60 mg, xs). The mixture was stirred at r.t for 24 h before filtered through Celite and washed with THF. LCMS showed product peak at 1.30 min (Iodide tray) and 1.37 min (ES<sup>+</sup>). Solvent was removed to afford crude product (12 mg). Purification by preparative HPLC *Gradient 8*, product peak at 15 min unfortunately failed and only provided crude (1.2 mg). Despite seeing 1 peak on HPLC and MS clean NMR could not be obtained. Unpromising

biological test ( $IC_{50} > 2$  mM) and reassessment by computational suggested inactivity of this type of compound result in abandon of this compound. No further attempt to prepare **13** was made. HRMS  $C_{30}H_{47}N_9O_6$ : calc 630.3728, found 630.3729, error: 0.15 ppm.

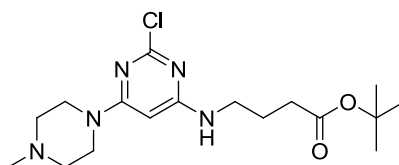
### 6.1.2. Heteroaromatic core

#### 2,4-Dichloro-6-(*N*-methylpiperazin-1-yl)pyrimidine, **47** <sup>164</sup>



2,4,6-Trichloropyrimidine (0.3 mL, 2.53 mmol) in EtOH (5.0 mL) was cooled to  $-25^{\circ}\text{C}$  and then  $\text{NEt}_3$  (0.39 mL, 2.80 mmol, 1.12 eq) was added to the solution. *N*-Methyl piperazine (0.28 mL, 2.50 mmol, 1.00 eq) was then added dropwise at  $-25^{\circ}\text{C}$ . The reaction mixture was stirred at  $-25^{\circ}\text{C}$  to  $0^{\circ}\text{C}$  over 4 h. A white precipitate was formed and then filtered off and washed with plenty of water and dried *in vacuo* to yield the product as a white solid (0.45g, 74%). Mpt:  $121-122^{\circ}\text{C}$ . IR ( $\nu$   $\text{cm}^{-1}$ ): 2843, 2937 (w,  $\text{sp}^3\text{-CH}$ ), 1575 (s,  $\text{ArC}=\text{C}$ ).  $^1\text{H-NMR}$  ( $\text{CDCl}_3$ , 600 MHz): 7.04 (s, 1H, Ar-**H**), 3.5-3.7 (br, 4H, 2  $\text{CH}_2\text{-CH}_2\text{-N}$ ), 2.36 (t, 4H,  $J$  5.0 Hz, 2  $\text{CH}_2\text{-N}(\text{CH}_3)$ ), 2.21 (s, 3H,  $\text{CH}_3$ ).  $^{13}\text{C-NMR}$  ( $\text{CDCl}_3$ , 150 MHz): 162.8 ( $\text{N-ArC}(\text{Cl})\text{-N}$ ), 159.1 ( $\text{ArC}(\text{N}((\text{CH}_2)(\text{CH}_2))\text{-ArN}$ ), 158.5 ( $\text{ArC}(\text{Cl})\text{-ArCH}$ ), 100.8 ( $\text{Ar-CH}$ ), 53.9 ( $\text{CH}_2\text{-CH}_2\text{-N}$ ), 46.2 ( $\text{CH}_2\text{-N}(\text{CH}_3)$ ), 45.5 ( $\text{CH}_3$ ). MS (EI)  $m/z$  (%): 248 (63) ( $^{37}\text{Cl}^{35}\text{Cl}$ ), 246 (100) ( $^{35}\text{Cl}^{35}\text{Cl}$ ). HRMS  $\text{C}_9\text{H}_{12}\text{Cl}_2\text{N}_4$ : calc. 246.04335, found 246.04336, error: 1.26 ppm.

#### *tert*-Butyl 4-(2-chloro-6-(4-methylpiperazin-1-yl)pyrimidin-4-ylamino)butanoate, **46**



To a solution of 2,4-dichloro-6-(*N*-methylpiperazin-1-yl)pyrimidine, **47** (136 mg, 0.55 mmol) in *n*-BuOH (4 mL) was added DIPEA (0.55 mL, 3.92 mmol, xs) and *tert*-butyl-4-aminobutanoate (158 mg, 0.99 mmol, 1.80 eq). The reaction mixture was heated under reflux for 26 h and white precipitate was formed. The white precipitate was filtered off and washed with EtOAc. The combined filtrate was then concentrated *in vacuo* and redissolved in  $\text{CH}_2\text{Cl}_2$ . The organic solution was then washed with  $\text{NaHCO}_3$  20% (2 x 20 mL), and the combined organic layers were dried ( $\text{MgSO}_4$ ). The solvent was removed *in vacuo*. The residues were purified by flash column chromatography ( $\text{CH}_2\text{Cl}_2$  with 0 to 2% MeOH) to give the product **46** as a light yellow oil (153 mg, 75%).  $R_f$ : 0.57 (10% MeOH in  $\text{CH}_2\text{Cl}_2$ ). IR ( $\nu$   $\text{cm}^{-1}$ ): 2978 (w,  $\text{sp}^3\text{-CH}$ ), 1722 (m,  $\text{C}=\text{O}$ ), 1576 (s,  $\text{ArC}=\text{C}$ ).  $^1\text{H-NMR}$  ( $\text{CDCl}_3$ , 600 MHz): 5.85 (s, 1H, Ar-**H**), 3.78-3.80 (br, 1H, N-**H**), 3.57-3.65 (br, 4H, 2  $\text{CH}_2\text{-CH}_2\text{-N}$  piperazine), 3.35-3.39 (m, 2H,  $\text{CH}_2\text{-COO}t\text{-Bu}$ ), 2.43-2.45 (br, 4H, 2  $\text{CH}_2\text{-N}(\text{CH}_3)$  piperazine), 2.31 (s, 3H, N- $\text{CH}_3$ ), 2.28 (t, 2H,  $J$  7.2 Hz,  $\text{COO}t\text{-Bu}$ ).

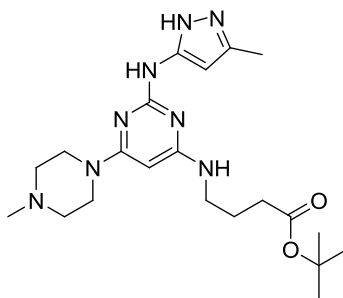


Bu-(CH<sub>2</sub>)<sub>2</sub>-CH<sub>2</sub>-NH), 1.80 (m, 2H, COO*t*-BuCH<sub>2</sub>-CH<sub>2</sub>-CH<sub>2</sub>NH), 1.41 (s, 9H, (CH<sub>3</sub>)<sub>3</sub>-C). <sup>13</sup>C-NMR (CDCl<sub>3</sub>, 150 MHz): 173.0 (C=O), 161.7 (NH-ArC-ArN), 161.2 ((ArC(N((CH<sub>2</sub>)(CH<sub>2</sub>))-ArN), 91.1(Ar-CH), 80.8 (C(CH<sub>3</sub>)<sub>3</sub>), 55.3 (CH<sub>2</sub>-CH<sub>2</sub>-N piperazine), 46.1 (CH<sub>2</sub>-N(CH<sub>3</sub>) piperazine), 44.1 (CH<sub>2</sub>-N(CH<sub>3</sub>) piperazine), 40.8 (CH<sub>2</sub>-COO*t*-Bu), 33.1 (COO*t*-Bu-(CH<sub>2</sub>)<sub>2</sub>-CH<sub>2</sub>-NH), 28.3 (CH<sub>3</sub>)<sub>3</sub>-C, 25.3 (COO*t*-BuCH<sub>2</sub>-CH<sub>2</sub>-CH<sub>2</sub>NH). MS (CI) *m/z* (%): 371 (19) (<sup>37</sup>Cl), 369 (56) (<sup>35</sup>Cl), 299 (28), 298 (26), 297 (14), 296 (77), 245 (26), 243 (75), 202 (23), 187 (100). HRMS C<sub>19</sub>H<sub>28</sub>ClN<sub>4</sub>O<sub>2</sub>: calc. 369.19260, found 369.19299, error: 1.06 ppm.

**(dba)Pd(*R*)-BINAP** <sup>135</sup>

To over-dried glassware which was then placed *in vacuo* and flushed with Argon several times, were added Pd<sub>2</sub>(dba)<sub>3</sub> (37.6 mg, 0.04 mmol) in dry benzene (3.5 mL) and (*R*)-BINAP (52 mg, 0.08 mmol, 2 eq). The purple solution was stirred at r.t for 2.5 h. The resulting dark orange solution was then filtered through Celite and the Celite was washed with toluene (15.0 mL). The solvent was removed *in vacuo*, and the oily residues were re-dissolved in diethyl ether (3.4 mL). The orange precipitate forming over 4 h was then collected, washed with diethyl ether and then dried *in vacuo* overnight to give the product catalyst (31.8 mg, 81%). The catalyst is stored inside a desiccator. <sup>1</sup>H-NMR (CDCl<sub>3</sub>, 400 MHz): 6.32-7.42 (m). <sup>31</sup>P-NMR (CDCl<sub>3</sub>, 162 MHz): 24.7-26 (br). (Identical by <sup>1</sup>H- and <sup>31</sup>P-NMR to the literature <sup>135</sup>)

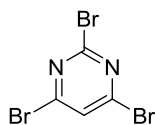
***tert*-Butyl-4-(2-(3-methyl-1H-pyrazol-5-ylamino)-6-(4-methylpiperazin-1-yl)pyrimidin-4-ylamino)butanoate, 3**



*tert*-Butyl 4-(2-chloro-6-(4-methylpiperazin-1-yl)pyrimidin-4-ylamino)butanoate, **46** (12.2 mg, 0.03 mmol) was dissolved in dry THF (3.5 mL). 3-Amino 5-methyl pyrazole (5.00 mg, 0.05 mmol, 1.70 eq) was added to the solution along with Cs<sub>2</sub>CO<sub>3</sub> (15.0 mg, 0.05 mmol, 1.60 eq) and the catalyst (dba)Pd(*R*)-BINAP) (5.00 mg, 5.00 μmol, 15 mol%). The resulting orange solution was then heated under microwave at 120°C for 10 min followed by 140°C for 30 min. LCMS confirmed strong product peak and no starting material remained. The solid was filtered off the reaction solution and washed with THF. The combined filtrate was concentrated *in vacuo* and the residues were dried *in vacuo*. Flash column chromatography (Chloroform with 1 to 10% MeOH) yielded the product **3** as a light yellow oil (11.2 mg, 80%). R<sub>f</sub>: 0.15 (10% MeOH in CH<sub>2</sub>Cl<sub>2</sub>). <sup>1</sup>H-NMR (CD<sub>3</sub>OD, 600 MHz): 6.33 (s, 1H, Ar-**H** pyrimidine), 5.28 (s, 1H, Ar-**H** pyrazole), 3.68-3.80 (br and m, 5H, (pyrimidine)-N-H-(pyrazole) and 2 CH<sub>2</sub>-CH<sub>2</sub>-N

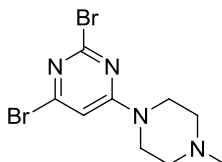
piperazine), 3.30 (t, 2H,  $J$  7.2 Hz, COO*t*-Bu-(CH<sub>2</sub>)<sub>2</sub>-CH<sub>2</sub>-NH), 3.10 (s, 1H, aliphatic N-H), 2.47-2.52 (m, 4H, 2 CH<sub>2</sub>-N(CH<sub>3</sub>) piperazine), 2.33 (s, 3H, N-CH<sub>3</sub> piperazine), 2.31 (t, 2H,  $J$  7.3 Hz, CH<sub>2</sub>-COO*t*-Bu), 2.11 (s, 3H, CH<sub>3</sub> pyrazole), 1.84-1.88 (quint, 2H,  $J$  7.2 Hz, COO*t*-BuCH<sub>2</sub>-CH<sub>2</sub>-CH<sub>2</sub>NH), 1.40 (s, 3H, (CH<sub>3</sub>)<sub>3</sub>-C). <sup>13</sup>C-NMR (CD<sub>3</sub>OD, 150 MHz): 174.8 (C=O), 162.7 (ArN-ArC(NH)-ArN pyrimidine), 162.5 ((ArC(N((CH<sub>2</sub>)(CH<sub>2</sub>))-ArN pyrimidine), 162.0 (ArCH-ArC-NH((CH<sub>2</sub>)<sub>3</sub>COO*t*-Bu)), 152.6 (ArC-NH pyrazole, ArC-CH<sub>3</sub> pyrazole), 90.1 (ArC-H pyrimidine), 81.5 (C(CH<sub>3</sub>)<sub>3</sub>), 79.6 (ArC-H pyrazole), 56.7 (CH<sub>2</sub>-CH<sub>2</sub>-N piperazine), 46.1 (CH<sub>2</sub>-N(CH<sub>3</sub>) piperazine), 41.6 (CH<sub>2</sub>-COO*t*-Bu), 34.0 (COO*t*-Bu-(CH<sub>2</sub>)<sub>2</sub>-CH<sub>2</sub>-NH), , 28.4 (CH<sub>3</sub>)<sub>3</sub>-C), 26.2 (COO*t*-BuCH<sub>2</sub>-CH<sub>2</sub>-CH<sub>2</sub>NH). MS (CI)  $m/z$  (%): 432 (37), 431 (100), 430 (26). HRMS C<sub>21</sub>H<sub>34</sub>N<sub>8</sub>O<sub>2</sub>: calc. 431.28830, found 431.28877, error: 1.09 ppm.

### 2,4,6-Tribromopyrimidine, **56** <sup>165</sup>



To an oven-dried, argon-filled flask, was added phosphorus (V) oxytribromide (5.00 g, 16.92 mmol, 4.0 eq), dry toluene (15.0 mL), barbituric acid (0.57 g, 4.46 mmol) and *N,N*-dimethylaniline (1.1 mL, 8.28 mmol, 1.9 eq). The mixture turned red and was heated under reflux for 3 h before cooling to r.t and then ice (20 g) was added. The mixture was then stirred to form a milky solution. The aqueous layer was separated and extracted with benzene (2 x 20 mL). The combined organic layers were dried (MgSO<sub>4</sub>), filtered and re-dissolved in MeOH (60 mL). The solvent was then removed *in vacuo* to give **56** as a yellow solid (0.96 g, 68%). Mpt: 91 – 95°C. IR (ν cm<sup>-1</sup>): 1504 (s, ArC=C). <sup>1</sup>H-NMR (CDCl<sub>3</sub>, 500 MHz): 7.74 (s, 1H, Ar-H). <sup>13</sup>C-NMR (CDCl<sub>3</sub>, 125 MHz): 153.1 (N-ArC(Br)-N), 150.8 (C-ArC(Br)-N), 128.2 (ArC-H). (Identical by <sup>1</sup>H- and <sup>13</sup>C-NMR to the literature <sup>144</sup>) MS (CI)  $m/z$  (%): 321(28), 319 (96) (<sup>81</sup>Br<sup>81</sup>Br<sup>79</sup>Br), 317 (100) (<sup>79</sup>Br<sup>79</sup>Br<sup>81</sup>Br), 315 (32).

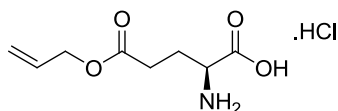
### 2,4-Dibromo-6-(*N*-methylpiperazin-1-yl)pyrimidine, **57**



2,4,6-Tribromopyrimidine, **56** (200 mg, 0.63 mmol) in MeOH (6.0 mL) was cooled to -30°C and then NEt<sub>3</sub> (0.2 mL, 1.43 mmol, 2.3 eq) was added to the solution. *N*-methyl piperazine (70 μL, 0.63 mmol, 1.0 eq) was then added dropwise at -30°C. The reaction mixture was stirred at -30°C to -5°C over 3 h. After that, the slightly yellow milky solution was allowed to gradually warm up to r.t and it turned transparent over time. The solvent was then removed *in vacuo*. Purification by flash column chromatography (CH<sub>2</sub>Cl<sub>2</sub> with 0 to 7% MeOH) yielded a white

solid which was washed with cold water to give **57** (118 mg, 56%).  $R_f$ : 0.29 (10% MeOH in  $\text{CH}_2\text{Cl}_2$ ) – as a salt with  $\text{HN}^+\text{Et}_3$ . Mpt: product decomposed at 257 – 260°C, not melted. IR ( $\nu \text{ cm}^{-1}$ ): 2939, 2805 (w,  $\text{sp}^3\text{-CH}$ ), 1566 (s,  $\text{ArC}=\text{C}$ ).  $^1\text{H-NMR}$  ( $\text{CDCl}_3$ , 600 MHz): 6.62 (s, 1H, **Ar-H**), 3.73-3.85 (br, 4H, 2  $\text{CH}_2\text{-CH}_2\text{-N}$ ), 2.50-2.60 (br, 4H, 2  $\text{CH}_2\text{-N}(\text{CH}_3)$ ), 2.36 (s, 3H,  $\text{CH}_3$ ).  $^{13}\text{C-NMR}$  ( $\text{CDCl}_3$ , 150 MHz): 162.1 ( $\text{ArC}(\text{N}((\text{CH}_2)(\text{CH}_2))\text{-ArN})$ ), 150.9 ( $\text{ArC}(\text{Br})\text{-ArCH}$ ), 104.4 ( $\text{Ar-CH}$ ), 54.3 ( $\text{CH}_2\text{-CH}_2\text{-N}$ ), 46 ( $\text{CH}_2\text{-N}(\text{CH}_3)$ ), 31.1 ( $\text{CH}_3$ ). MS (EI) (%):  $m/z$ , 338 (47) ( $^{81}\text{Br}^{81}\text{Br}$ ), 336 (100) ( $^{81}\text{Br}^{79}\text{Br}$ ), 334 (50) ( $^{79}\text{Br}^{79}\text{Br}$ ). HRMS  $\text{C}_9\text{H}_{12}\text{Br}_2\text{N}_4$ : calc. 333.94232, found 333.94255, error: 0.69 ppm.

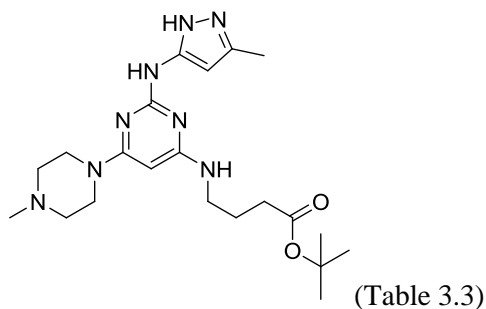
**(S)-2-amino-(5-allyloxy)-5-oxo-pentanoic acid. HCl salt, 60** <sup>166</sup>



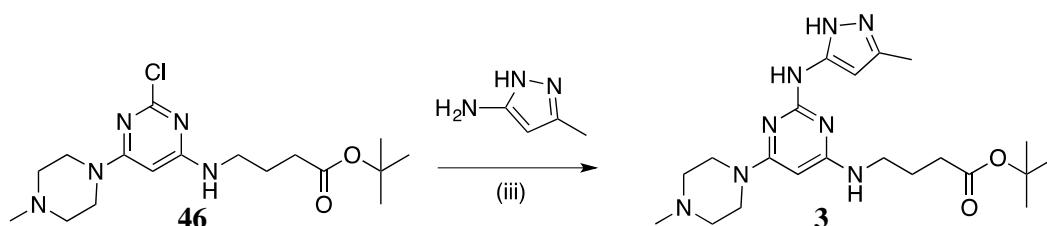
Commercially available L-glutamic acid, **59** (10.60 g, 71.3 mmol) and  $\text{TMSCl}$  (23 mL, xs) was added to a solution of allyl alcohol (250 mL, xs). The reaction mixture was then stirred at 0°C for 20 min then at r.t for 24 h. Ice-cold diethyl ether was then added (800 mL) and a white precipitate formed. The solution was stirred for 1 h to obtain more precipitate. The precipitate was then filtered and washed with cold diethyl ether (2 x 100 mL) and dried *in vacuo* to give the product **60** as the HCl salt (6.73 g, 42%). Mpt: 119-123°C. IR ( $\nu \text{ cm}^{-1}$ ): 3154(m, NH), 2974, 2907 (br, m, -OH and  $\text{sp}^3\text{-CH}$ ), 1739 (s,  $\text{C}=\text{O}$  ester), 1722 (s,  $\text{C}=\text{O}$  acid), 1610 (m,  $\text{C}=\text{C}$ ).  $^1\text{H-NMR}$  ( $(\text{CD}_3)_2\text{SO}$ , 500 MHz): 8.59 (br, 2H,  $\text{NH}_2$ ), 5.91 (m, 1H,  $\text{CH}_2=\text{CH}$ ), 5.29 (d,  $J$  16.2 Hz, 1H,  $\text{CH}_2=\text{CH}$ ), 5.20 (d,  $J$  10.8 Hz, 1H,  $\text{CH}_2=\text{CH}$ ), 4.53 (br, 2H,  $\text{CH}_2\text{O}$ ), 3.89 (br, 1H, **H-C**( $\alpha$ )), 2.62 (m, 2H, **HC**( $\gamma$ )), 2.05-2.11 (m, 2H, **H-C**( $\beta$ )).  $^1\text{H-NMR}$  ( $\text{CD}_3\text{OD}$ , 500 MHz): 5.91-5.97 (m, 1H,  $\text{COOH}$ ), 5.32 (dd, 1H,  $J$  17.2 and 1.5 Hz,  $\text{CH}_2=\text{CH}$ ), 5.22 (dd,  $J$  12.9 and 1.3 Hz, 1H,  $\text{CH}_2=\text{CH}$ ), 4.85 (br, 3H,  $\text{NH}_3$ ), 4.60 (dd, 2H,  $J$  7.7 and 0.9 Hz,  $\text{CH}_2\text{O}$ ), 4.05 (t, 2H,  $J$  7.0 Hz, **H-C**( $\alpha$ )), 2.60-2.66 (m, 2H, **HC**( $\gamma$ )), 2.01-2.28 (m, 2H, **H-C**( $\beta$ )).  $^{13}\text{C-NMR}$  ( $\text{CD}_3\text{OD}$ , 125 MHz): 173.3 ( $\text{C}=\text{O}(\text{OAllyl})$ ), 171.3 ( $\text{C}=\text{O}(\text{OH})$ ), 133.5 ( $\text{CH}_2=\text{CH}$ ), 118.5 ( $\text{CH}_2=\text{CH}$ ), 66.5 ( $\text{CH}_2\text{O}$ ), 53.1 (**H-C**( $\alpha$ )), 30.5 (**HC**( $\gamma$ )), 26.6 (**H-C**( $\beta$ )). (Identical by  $^1\text{H}$ - and  $^{13}\text{C}$ -NMR to the literature <sup>142</sup>) MS (CI)  $m/z$ , (%): 189 (22), 188 (100). HRMS  $\text{C}_8\text{H}_{13}\text{NO}_4$ : calc. 188.09228, found 188.09261, error: 1.75 ppm.

### Failed synthetic attempts

*tert*-Butyl 4-(2-(3-methyl-1H-pyrazol-5-ylamino)-6-(4-methylpiperazin-1-yl)pyrimidin-4-ylamino)butanoate, **3**



**Table 3.3**



(iii) Pd catalyst, ligand, base, solvent

Entry	Catalyst and ligand, mol%	Base	Solvent	Time (h)	Temperature	Yield %
1	(dba)Pd( <i>R</i> -BINAP), 1.3	Cs <sub>2</sub> CO <sub>3</sub>	dioxane	25	thermal heat, 100°C	×
2	Pd <sub>2</sub> (dba) <sub>3</sub> , DavePhos, 3.8	NaO <i>t</i> -Bu	dioxane	17	thermal heat, 100°C	×
3	(dba)Pd( <i>R</i> -BINAP), 4	NaO <i>t</i> -Bu	dioxane	40	thermal heat, 95°C	×
4	Pd(PPh <sub>3</sub> ) <sub>4</sub> , <i>R</i> -BINAP, 8	NaO <i>t</i> -Bu	toluene	27	thermal heat, reflux	×
5	(dba)Pd( <i>R</i> -BINAP), 5	Cs <sub>2</sub> CO <sub>3</sub>	THF	1.6	microw, 140°C	6% isolated product
6	(dba)Pd( <i>R</i> -BINAP), 15	Cs <sub>2</sub> CO <sub>3</sub>	THF	0.6	microw, 120 – 140°C	80% isolated product

#### *Entry 1*

*tert*-Butyl 4-(2-chloro-6-(4-methylpiperazin-1-yl)pyrimidin-4-ylamino)butanoate, **46** (89.3 mg, 0.24 mmol) was dissolved in dry dioxane (3.0 mL). 3-Amino 5-methyl pyrazole (28.0 mg, 0.28 mmol, 1.17 eq) was added to the solution along with Cs<sub>2</sub>CO<sub>3</sub> (113 mg, 0.34 mmol, 1.42 eq) and the catalyst (dba)Pd(*R*-BINAP) (3.0 mg, 3 μmol, 1.3 mol%). The resulting solution was then

heated at 100°C for 25 h. LCMS confirmed no product peak and starting material. Reaction failed.

*Entry 2*

*tert*-Butyl 4-(2-chloro-6-(4-methylpiperazin-1-yl)pyrimidin-4-ylamino)butanoate, **46** (10.6 mg, 28.7  $\mu$ mol) was dissolved in dry dioxane (1.5 mL). 3-Amino 5-methyl pyrazole (2.60 mg, 26.8  $\mu$ mol, 0.96 eq) was added to the solution along with NaO*t*-Bu (2.80 mg, 29.4  $\mu$ mol, 1.02 eq) and the catalysts (dba)<sub>2</sub>Pd(dba)<sub>3</sub> (0.9 mg, 0.98  $\mu$ mol, 3.8 mol%) and DavePhos (2.0 mg, 7.11  $\mu$ mol). The resulting solution was then heated at 100°C for 17 h. LCMS confirmed no product peak. Reaction failed.

*Entry 3*

*tert*-Butyl 4-(2-chloro-6-(4-methylpiperazin-1-yl)pyrimidin-4-ylamino)butanoate, **46** (34.0 mg, 92.0  $\mu$ mol) was dissolved in dry dioxane (3.5 mL). 3-Amino 5-methyl pyrazole (13.0 mg, 0.13 mmol, 1.41 eq) was added to the solution along with NaO*t*-Bu (16.0 mg, 0.17 mmol, 1.80 eq) and the catalysts (dba)<sub>2</sub>Pd(dba)<sub>3</sub> (3.50 mg, 3.80  $\mu$ mol, 4.0 mol%) and ((*R*)-BINAP (7.0 mg, 11.0  $\mu$ mol). The resulting solution was then heated at 95°C for 40 h. LCMS confirmed no product peak. Reaction failed.

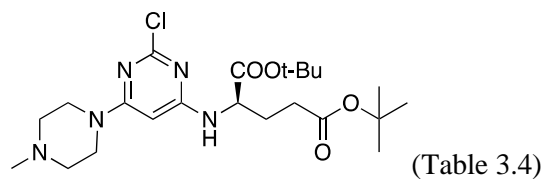
*Entry 4*

*tert*-Butyl 4-(2-chloro-6-(4-methylpiperazin-1-yl)pyrimidin-4-ylamino)butanoate, **46** (17.5 mg, 47.3  $\mu$ mol) was dissolved in dry toluene (4.2 mL). 3-Amino 5-methyl pyrazole (5.10 mg, 52.5  $\mu$ mol, 1.11 eq) was added to the solution along with NaO*t*-Bu (12.0 mg, 0.13 mmol, 2.64 eq) and the catalysts Pd(PPh<sub>3</sub>)<sub>4</sub> (6.0 mg, 5.19  $\mu$ mol, 8 mol%) and *R*-BINAP (2.95 mg, 4.73  $\mu$ mol). The resulting solution was then heated under reflux for 27 h. LCMS confirmed no product peak. Reaction failed.

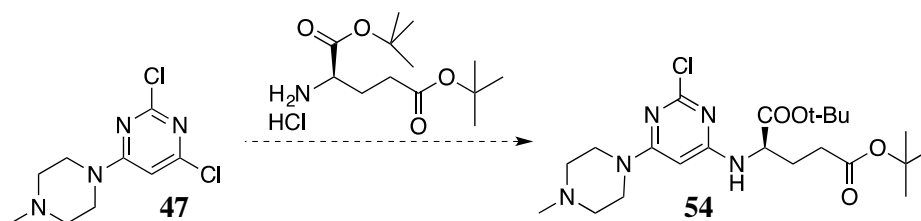
*Entry 5*

*tert*-Butyl 4-(2-chloro-6-(4-methylpiperazin-1-yl)pyrimidin-4-ylamino)butanoate, **46** (66.0 mg, 0.18 mmol) was dissolved in dry THF (4.0 mL). 3-Amino 5-methyl pyrazole (27.0 mg, 0.27 mmol, 1.52 eq) was added to the solution along with Cs<sub>2</sub>CO<sub>3</sub> (87.0 mg, 0.17 mmol, 1.52 eq) and the catalyst (dba)Pd((*R*)-BINAP) (8.00 mg, 8.00  $\mu$ mol, 5 mol%). The resulting solution was then heated under microwave at 140°C for 1.60 h. LCMS confirmed product peak and starting material. The solid was filtered off the reaction solution and washed with THF. The combined filtrate was concentrated *in vacuo* and the residues were dried *in vacuo*. Flash column chromatography (Chloroform with 1 to 3% MeOH) yielded the product **3** as a light yellow oil (4.70 mg, 6%).

**Di-*tert*-butyl (2-chloro-6-(4-methylpiperazin-1-yl)pyrimidin-4-yl)-D-glutamate, 54**



**Table 3.4**



Entry	Amine (1.5-2 eq)	Base (+catalyst) (xs)	Solvent	Temperature	Time (h)	Yield %
1	salt	DIPEA	<i>n</i> -BuOH and EtOH	reflux	7	6% of crude <b>54</b>
2	salt	DIPEA	<i>n</i> -BuOH	reflux	40	X
3	salt	DIPEA	THF	reflux	4	X
4	salt	Et <sub>3</sub> N	EtOH	reflux	24	X
5	free	DIPEA	<i>t</i> -BuOH	reflux	24	X
6	free	DIPEA	<i>n</i> -BuOH and EtOH	reflux	72	X
7	salt	Cs <sub>2</sub> CO <sub>3</sub> (+(dba)Pd(( <i>R</i> )- BINAP))	toluene	microw, 140°C	25	X
8	salt	Cs <sub>2</sub> CO <sub>3</sub> (+(dba)Pd(( <i>R</i> )- BINAP))	THF	microw, 140°C	30	X

**Free-form di-*tert*-butyl D-glutamate**

Di-*tert*-butyl D-glutamate hydrochloride (178 mg, 0.60 mmol) was dissolved in water (pH = 4.7). To the solution, was added saturated NaHCO<sub>3</sub> solution and NaOH 10M until the pH reached 9.5. The solution was extracted by chloroform (3 x 40 mL). The combined extract was dried (MgSO<sub>4</sub>) and concentrated *in vacuo* to give the free-form di-*tert*-butyl D-glutamate as a colourless oil (146 mg, 94%). This was used directly in further experiments.

#### Entry 1

To a solution of 2,4-dichloro-6-(*N*-methylpiperazin-1-yl)pyrimidine, **47** (69.0 mg, 0.28 mmol) in *n*-BuOH (3 mL) and EtOH (5 mL), were added DIPEA (0.25 mL, 1.78 mmol, xs) and di-*tert*-butyl D-glutamate hydrochloride (134 mg, 0.45 mmol, 1.60 eq). The reaction mixture was heated under reflux for 7 h. The reaction mixture was then concentrated *in vacuo* and purified by flash column chromatography (CH<sub>2</sub>Cl<sub>2</sub> with 0 to 2% MeOH) to give a crude mixture containing product **54** (7.8 mg, 6%). <sup>1</sup>H-NMR showed that only 20% of the crude was the product: <sup>1</sup>H-NMR (CDCl<sub>3</sub>, 500 MHz): 6.40 (s, 1H, Ar-**H**), 3.30-3.40 (br, m, 4H, 2 CH<sub>2</sub>-CH<sub>2</sub>-N piperazine), 2.68-2.77 (br, m, 4H, 2 CH<sub>2</sub>-N(CH<sub>3</sub>) piperazine), 2.48 (s, 3H, N-CH<sub>3</sub>), 2.40-2.45 (br, m, 1H, HN(C\***H**), underneath other peaks), 2.15-2.17 (m, 2H, -C\*(H)-CH<sub>2</sub>-CH<sub>2</sub>-COO*t*-Bu) 2.00-2.05 (m, 2H, -C\*(H)-CH<sub>2</sub>-CH<sub>2</sub>), 1.50 (s, 18H, (CH<sub>3</sub>)<sub>3</sub>-C, underneath other peaks).

#### Entry 2

To a solution of 2,4-dichloro-6-(*N*-methylpiperazin-1-yl)pyrimidine, **47** (107 mg, 0.43 mmol) in *n*-BuOH (5 mL), were added DIPEA (0.50 mL, 3.59 mmol, xs) and di-*tert*-butyl D-glutamate hydrochloride (209 mg, 0.71 mmol, 1.65 eq). The reaction mixture was heated under reflux for 40 h. LCMS did not detect any product.

#### Entry 3

To a solution of 2,4-dichloro-6-(*N*-methylpiperazin-1-yl)pyrimidine, **47** (21.4 mg, 86.6 μmol) THF (2 mL), were added DIPEA (0.30 mL, 2.15 mmol, xs) and di-*tert*-butyl D-glutamate hydrochloride (43.0 mg, 0.14 mmol, 1.62 eq). The reaction mixture was heated under reflux for 4 h. LCMS did not detect any product.

#### Entry 4

To a solution of 2,4-dichloro-6-(*N*-methylpiperazin-1-yl)pyrimidine, **47** (53.5 mg, 0.22 mmol) in EtOH (4 mL), were added Et<sub>3</sub>N (0.60 mL, 4.26 mmol, xs) and di-*tert*-butyl D-glutamate hydrochloride (100 mg, 0.34 mmol, 1.55 eq). The reaction mixture was heated under reflux for 24 h. LCMS did not detect any product.

#### Entry 5

To a solution of 2,4-dichloro-6-(*N*-methylpiperazin-1-yl)pyrimidine, **47** (72.0 mg, 0.29 mmol) in *t*-BuOH (5 mL), were added DIPEA (0.30 mL, 2.15 mmol, xs) and free-form di-*tert*-butyl D-glutamate (110 mg, 0.42 mmol, 1.45 eq). The reaction mixture was heated under reflux for 24 h. LCMS did not detect any product.

#### Entry 6

To a solution of 2,4-dichloro-6-(*N*-methylpiperazin-1-yl)pyrimidine, **47** (56.0 mg, 0.23 mmol) in *n*-BuOH (2.5 mL) and EtOH (1.6 mL), were added DIPEA (0.20 mL, 1.43 mmol, xs) and free-form di-*tert*-butyl D-glutamate (85.4 mg, 0.33 mmol, 1.43 eq). The reaction mixture was heated under reflux for 72 h. LCMS did not detect any product.

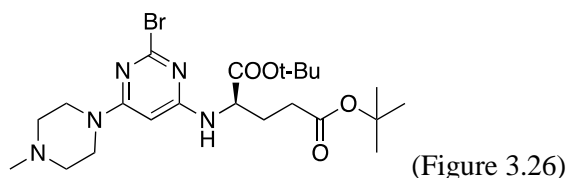
#### Entry 7

To a solution of 2,4-dichloro-6-(*N*-methylpiperazin-1-yl)pyrimidine, **47** (52.5 mg, 0.21 mmol) in toluene (3.0 mL), were added di-*tert*-butyl D-glutamate hydrochloride (75.0 mg, 0.25 mmol, 1.19 eq) along with Cs<sub>2</sub>CO<sub>3</sub> (81.0 mg, 0.25 mmol, 1.19 eq) and the catalysts (dba)<sub>2</sub>Pd(dba)<sub>3</sub> (45 mg, 0.05 mmol, 22 mol%) and ((*R*)-BINAP (44.5 mg, 0.07 mmol). The reaction mixture was heated under microwave at 140°C for 25 min. LCMS did not detect any product.

#### Entry 8

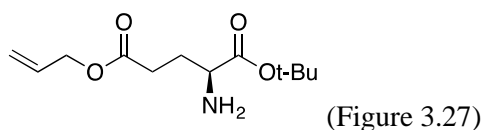
To a solution of 2,4-dichloro-6-(*N*-methylpiperazin-1-yl)pyrimidine, **47** (51.0 mg, 0.21 mmol) in THF (3.5 mL), were added di-*tert*-butyl D-glutamate hydrochloride (88.5 mg, 0.30 mmol, 1.42 eq) along with Cs<sub>2</sub>CO<sub>3</sub> (99.7 mg, 0.31 mmol, 1.42 eq) and the catalysts (dba)<sub>2</sub>Pd(dba)<sub>3</sub> (29 mg, 0.03 mmol, 14 mol%) and ((*R*)-BINAP (40.0 mg, 0.06 mmol). The reaction mixture was heated under microwave at 140°C for 30 min. LCMS did not detect any product.

#### Di-*tert*-butyl (2-bromo-6-(4-methylpiperazin-1-yl)pyrimidin-4-yl)-D-glutamate, **58**



To a solution of 2,4-dibromo-6-(*N*-methylpiperazin-1-yl)pyrimidine, **57** (50.8 mg, 0.15 mmol) in *n*-BuOH (5.3 mL) at r.t, were added di-*tert*-butyl D-glutamate hydrochloride (53.0 mg, 0.18 mmol, 1.20 eq) and DIPEA (26 μL, 0.19 mmol, 1.30 eq). The reaction mixture was heated under reflux over 5 h and monitored by TLC. After 5 h, only starting materials were detected. Further amounts of di-*tert*-butyl D-glutamate hydrochloride (33.0 mg, 0.1 mmol, 0.70 eq) and DIPEA (50 μL, 0.19 mmol, xs) were added. The reaction mixture was heated under reflux for 15 h. LCMS did not detect any signals of the product.

#### 5-Allyl 1-(*tert*-butyl) *L*-glutamate, **61**

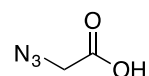




To a solution of (*S*)-2-amino-(5-allyloxy)-5-oxo-pentanoic acid. HCl salt, **60** (6.26 g, 28.0 mmol) in perchloric acid 70% (17 mL) at 0°C, *tert*-butyl alcohol (50 mL, 0.37 mol, 13.0 eq) was added dropwise over 1 h. The reaction was allowed to gradually warm up to r.t whilst stirring for 48 h. After that, the reaction mixture was cooled to 0°C and NaOH (10 M, 87 mL) was added slowly to change the pH of the solution to 10. The reaction mixture was then extracted with EtOAc (2 x 80 mL). The combined extract was dried (MgSO<sub>4</sub>). TCL (10% MeOH in CH<sub>2</sub>Cl<sub>2</sub>) did not detect the formation of the product.

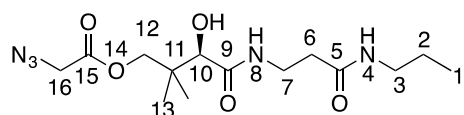
## 6.2. Approach B syntheses

### 2-Azidoacetic acid, **63** <sup>167</sup>



To a solution of 2-bromoacetic acid (5.0 g, 36.0 mmol) in water (25 mL) at 0°C, was added sodium azide (10.0 g, 153.8 mmol, xs). The reaction mixture was stirred from 0°C to r.t for 24 h. After that, a solution of HCl 37% was added (15 mL) until the colour of the mixture turned from peach to colourless. The combined organic layers were extracted with Et<sub>2</sub>O (3 x 100 mL) then dried (Na<sub>2</sub>SO<sub>4</sub>). The solvent was removed *in vacuo* to afford the product as a colourless liquid (3.65 g, quantitative). IR ( $\nu_{\max}$ , cm<sup>-1</sup>): 3100 (br, OH), 2918 (m, sp<sup>3</sup>-CH), 2105 (s, azide), 1716 (s, C=O acid). <sup>1</sup>H-NMR (D<sub>2</sub>O, 600 MHz): 4.10 (s, 2H, CH<sub>2</sub>). (Identical by <sup>1</sup>H-NMR to the literature <sup>147</sup>) <sup>13</sup>C-NMR (D<sub>2</sub>O, 150 MHz): 173.5 (C=O), 50.8 (CH<sub>2</sub>). MS (EI) *m/z* (%): 101 (100). HRMS C<sub>2</sub>H<sub>3</sub>N<sub>3</sub>O<sub>2</sub>: calc. 101.0220, found 101.0215, error: 0.46 ppm.

### (*R*)-3-hydroxy-2,2-dimethyl-4-oxo-4-((3-oxo-3-(propylamino)propyl)amino)butyl 2-azidoacetate, **64**



#### Procedure using DCC and DMAP

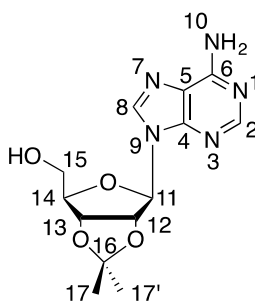
To a colourless solution of 2-azidoacetic acid, **63** (152 mg, 1.50 mmol, 1.50 eq) in THF (4.0 mL) at 0°C, was added DCC (385 mg, 1.80 mmol, 1.80 eq) to form a cloudy solution. DMAP (20 mg, 0.16 mmol, 0.16 eq) was added followed by (*R*)-2,4-dihydroxy-3,3-dimethyl-*N*-(3-oxo-3-(propylamino)propyl)butanamide, **20** (260 mg, 1.00 mmol) in THF (2.7 mL). A further 4.0 mL of THF was added and the reaction mixture was stirred from 0°C to r.t for 40 min before heating at 45°C whilst stirring for 20 h. The solvent was then removed *in vacuo*. Purification by flash column chromatography (Petroleum Ether to Petroleum Ether : EtOAc = 1:1 then 1:2 then 100% EtOAc to 0.5 to 1% MeOH in EtOAc) gave the product **64** as a white solid (150 mg, 44%). R<sub>f</sub>: 0.20 (EtOAc).

*Procedure using PyAOP and HOAt*

To a colourless solution of 2-azidoacetic acid, **63** (111 mg, 1.09 mmol, 1.02 eq) in DMF (9.0 mL) at 0°C, were added PyAOP (1.03 g mg, 1.98 mmol, 1.85 eq), HOAt (206 mg, 1.51 mmol, 1.41 eq) and DIPEA (0.20 mL, 1.15 mmol, 1.07 eq) followed by (*R*)-2,4-dihydroxy-3,3-dimethyl-*N*-(3-oxo-3(propylamino)propyl)butanamide, **20** (280 mg, 1.07 mmol) in DMF (10 mL). The reaction mixture was stirred from 0°C to r.t for 18 h. The solvent was then removed *in vacuo* and Petroleum Ether (10 mL) and CH<sub>2</sub>Cl<sub>2</sub> (50 mL) were added. After filtration, the filtrate was concentrated *in vacuo*. Purification by flash column chromatography (Petroleum Ether : EtOAc = 3:1 then 2:1 then 100% EtOAc to 0.2 to 1% MeOH in EtOAc) gave the product **64** as a white solid (209 mg, 57%).

IR ( $\nu_{\max}$ , cm<sup>-1</sup>): 3319 (m, br, NH, OH), 2929 (m, sp<sup>3</sup>-CH), 2105 (s, azide), 1741 (s, C=O ester), 1639 (s, C=ONH), 1527 (s, CONH). <sup>1</sup>H-NMR (CDCl<sub>3</sub>, 600 MHz): 7.31 (t, br, 1H, N-H, H-8), 6.02 (s, br, 1H, N-H, H-4), 4.20 (d, 1H, *J* 10.92 Hz, H-12), 3.99 (d, 1H, *J* 10.92 Hz, H-12), 3.93 (s, 1H, CH, H-10), 3.92 (s, 2H, CH<sub>2</sub>, H-16), 3.52-3.55 (m, 2H, CH<sub>2</sub>, H-7), 3.16-3.19 (m, 2H, CH<sub>2</sub>, H-3), 2.41 (t, 2H, CH<sub>2</sub>, *J* 5.16 Hz, H-6), 1.47-1.53 (m, 2H, CH<sub>2</sub>, H-2), 1.03 (s, 3H, CH<sub>3</sub>, H-13), 0.98 (s, 3H, CH<sub>3</sub>, H-13), 0.94 (t, 3H, CH<sub>3</sub>, *J* 7.3 Hz, H-1). <sup>13</sup>C-NMR (CDCl<sub>3</sub>, 150 MHz): 172.4 (C=O, C-9), 172.3 (C=O, C-15), 171.6 (C=O, C-5), 71.7 (CH<sub>2</sub>, C-12), 75.1 (CH, C-10), 50.6 (CH<sub>2</sub>, C-16), 41.6 (CH<sub>2</sub>, C-3), 38.4 (C, C-11), 35.6 (CH<sub>2</sub>, C-6), 35.4 (CH<sub>2</sub>, C-7), 22.8 (CH<sub>2</sub>, C-2), 20.9 (CH<sub>3</sub>, C-13), 20.8 (CH<sub>3</sub>, C-13), 11.5 (CH<sub>2</sub>, C-1). MS (EI) *m/z* (%): 344 (100), 345 (20). HRMS C<sub>14</sub>H<sub>26</sub>N<sub>5</sub>O<sub>5</sub>: calc. 344.1928, found 344.1931, error: 0.20 ppm.

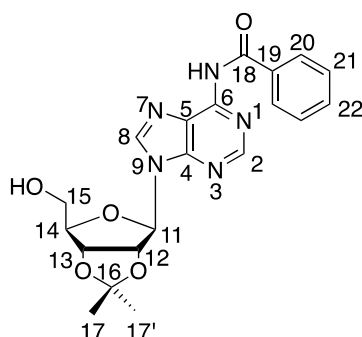
**((3*aR*,4*R*,6*R*,6*aR*)-6-(6-amino-9*H*-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)methanol, **66**<sup>168</sup>**



To a mixture of adenosine (3.00 g, 11.22 mmol) in acetone (300 mL) at 0°C, was added *p*-toluenesulfonic acid (21.8 g, 113 mmol, xs) in two equal portions. The mixture was stirred at 0°C to 11°C for 20 h. Saturated NaHCO<sub>3</sub> solution (300 mL) was then added to the mixture whilst stirring and acetone was removed *in vacuo*. NaOH 10 M was added to bring pH to 11. The combined organic layers were extracted with EtOAc (3 x 200 mL) then dried (NaSO<sub>4</sub>) and concentrated *in vacuo* to afford the product as a white solid (3.40 g, 99%). IR ( $\nu_{\max}$ , cm<sup>-1</sup>): 3267, 3106, 3061 (br, w, NH<sub>2</sub>, OH), 2986, 2937 (w, sp<sup>3</sup>-CH), 1682, 1570 (m, adenine C=C, C=N).

$^1\text{H-NMR}$  ( $(\text{CD}_3)_2\text{SO}$ , 600 MHz): 8.35 (s, 1H, Ar-H, H-8), 8.15 (s, 1H, Ar-H, H-2), 7.34 (br, 2H, N-H), 6.12 (d, 1H, CH,  $J$  2.83 Hz, H-11), 5.34 (dd, 1H, CH,  $J$  6.00 Hz and 2.77 Hz, H-12), 5.25 (t, 1H, O-H,  $J$  5.80 Hz), 4.96 (dd, 1H, CH,  $J$  6.29 Hz and 2.42 Hz, H-13), 4.20-4.22 (m, 1H, CH, H-14), 3.49-3.58 (m, 2H, CH<sub>2</sub>, H-15), 1.54 (s, 3H, CH<sub>3</sub>, H-17'), 1.32 (s, 3H, CH<sub>3</sub>, H-17). (Identical by  $^1\text{H-NMR}$  to the literature <sup>148</sup>)  $^{13}\text{C-NMR}$  ( $(\text{CD}_3)_2\text{SO}$ , 150 MHz): 156.2 (Ar-C, C-6), 152.8 (Ar-C, C-2), 148.8 (Ar-C, C-4), 139.7 (Ar-C, C-8), 119.2 (Ar-C, C-5), 113.1 (C, C-16), 89.7 (CH, C-11), 86.5 (CH, C-14), 83.4 (CH, C-12), 81.4 (CH, C-13), 61.6 (CH<sub>2</sub>, C-15), 27.2 (CH<sub>3</sub>, C-17'), 25.2 (CH<sub>3</sub>, C-17). MS (ESI)  $m/z$  (%): 308 (100), 309 (25). HRMS C<sub>13</sub>H<sub>18</sub>N<sub>5</sub>O<sub>4</sub>: calc. 308.1359, found 308.1351, error: 2.60 ppm.

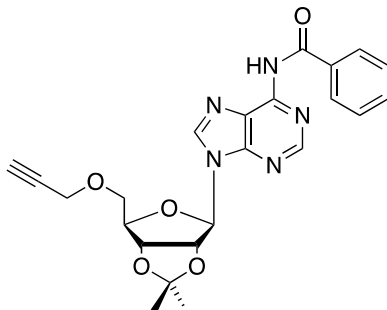
***N*-9-((3*aR*,4*R*,6*R*,6*aR*)-6-(hydroxymethyl)-2,2-dimethyltetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)-9*H*-purin-6-yl)benzamide, **67** <sup>169</sup>**



((3*aR*,4*R*,6*R*,6*aR*)-6-(6-Amino-9*H*-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)methanol, **66** (500 mg, 1.62 mmol) was added to pyridine (10 mL) and the solution was co-evaporated before pyridine (9.0 mL) was added back to the mixture. TMSCl (1.05 mL, 8.27 mmol, xs) was added whilst stirring at r.t for 30 min followed by BzCl (296 mg, 2.11 mmol, 1.30 eq). The mixture was stirred at r.t for 3.5 h before cooling down to 0°C. Water (1.65 mL) was added whilst stirring for 10 min then ammonia 28% solution (3.3 mL) was added. The mixture was stirred at r.t for 45 min then evaporated. Purification by flash column chromatography (0 to 2.5% MeOH in CH<sub>2</sub>Cl<sub>2</sub>) gave the product **67** as a white solid (424 mg, 64 %). R<sub>f</sub>: 0.40 (10% MeOH in CH<sub>2</sub>Cl<sub>2</sub>). IR ( $\nu_{\text{max}}$ , cm<sup>-1</sup>): 3351, 3180, 3061 (br, w, NH<sub>2</sub>, OH), 2986, 2923, 2853 (w, sp<sup>3</sup>-CH), 1707 (s, C=ONH), 1602, 1588, 1524 (m, adenine C=C, C=N and Ph(C=C)).  $^1\text{H-NMR}$  ( $(\text{CD}_3)_2\text{SO}$ , 600 MHz): 11.24 (s, 1H, N-H), 8.77 (s, 1H, Ar-H, H-8), 8.68 (s, 1H, Ar-H, H-2), 8.05 (d, 2H, Ar-H,  $J$  7.28 Hz, H-20), 7.66 (t, 1H, Ar-H,  $J$  7.67 Hz, H-22), 7.56 (t, 2H, Ar-H,  $J$  7.70 Hz, H-21), 6.28 (d, 1H, CH,  $J$  2.34 Hz, H-11), 5.45 (dd, 1H, CH,  $J$  6.20 Hz and 2.26 Hz, H-12), 5.16 (t, 1H, O-H,  $J$  5.80 Hz), 5.01 (dd, 1H, CH,  $J$  6.20 Hz and 2.26 Hz, H-13), 4.27 (m, 1H, CH, H-14), 3.55-3.56 (m, 2H, CH<sub>2</sub>, H-15), 1.56 (s, 3H, CH<sub>3</sub>, H-17'), 1.35 (s, 3H, CH<sub>3</sub>, H-17).  $^{13}\text{C-NMR}$  ( $(\text{CD}_3)_2\text{SO}$ , 150 MHz): 165.6 (C=O, C-18), 151.8 (Ar-C, C-2), 150.5 (Ar-C, C-4), 143.2 (Ar-C, C-8), 134.3 (Ar-C, C-6), 133.3 (Ar-C, C-19), 132.5 (Ar-C, C-22), 128.6 (Ar-C, C-20), 128.2 (Ar-C, C-21), 119.2 (Ar-C, C-5), 113.1 (C, C-16), 89.3 (CH, C-11), 86.9 (CH, C-14), 83.7 (CH, C-12), 81.8 (CH, C-13), 61.5 (CH<sub>2</sub>, C-15), 27.2

(CH<sub>3</sub>, C-17'), 25.2 (CH<sub>3</sub>, C-17). (Identical by <sup>1</sup>H- and <sup>13</sup>C- NMR to the literature <sup>149</sup>) HRMS C<sub>20</sub>H<sub>22</sub>N<sub>5</sub>O<sub>5</sub>: calc. 412.1621, found 412.1628, error: 1.70 ppm.

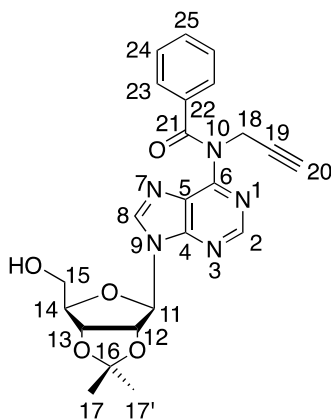
***N***-9-((3*aR*,4*R*,6*R*,6*aR*)-2,2-dimethyl-6-((prop-2-yn-1-yloxy)methyl)tetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)-9*H*-purin-6-yl)benzamide, **68** <sup>149</sup> (Attempted)



(Figure 4.3)

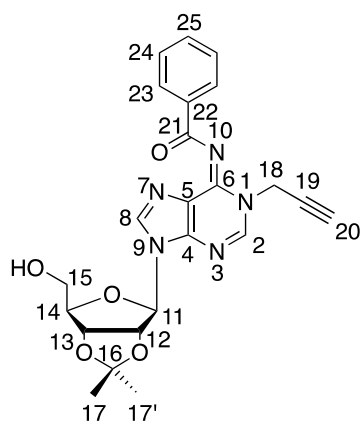
To a solution of **67** (974 mg, 2.36 mmol) in THF (20 mL) at 0°C, was added NaH (105 mg, 2.62 mmol, 1.10 eq) slowly. The reaction mixture was stirred at 0°C for 1 h then TBAI (290 mg, 0.78 mmol, 0.33 eq) and THF (20 mL) were added in followed by propargyl bromide 80% (390 mg, 1.30 eq). The reaction mixture was stirred at r.t for 24 h. Purification by flash column chromatography (0 to 4% MeOH in CH<sub>2</sub>Cl<sub>2</sub>) gave the product **69** (240 mg) and **70** (440 mg) as white foams.

***N***-9-((3*aR*,4*R*,6*R*,6*aR*)-6-(hydroxymethyl)-2,2-dimethyltetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)-9*H*-purin-6-yl)-*N*-(prop-2-yn-1-yl)benzamide, **69**



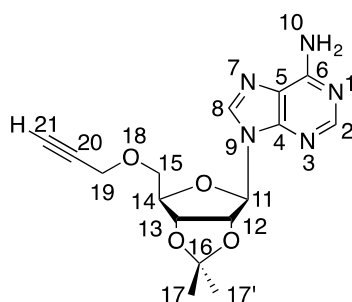
<sup>1</sup>H-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 600 MHz): 8.68 (s, 1H, Ar-**H**, H-8), 8.62 (s, 1H, Ar-**H**, H-2), 7.36-7.37 (m, 3H, Ar-**H**, H-23 and H-25), 7.26 (m, 2H, Ar-**H**, H-24), 6.19 (d, 1H, **CH**, *J* 2.90 Hz, H-11), 5.35 (dd, 1H, **CH**, *J* 6.20 Hz and 2.90 Hz, H-12), 5.13 (t, 1H, O-**H**, *J* 5.30 Hz), 5.07 (d, 2H, **CH**<sub>2</sub>, 2.50 Hz, H-18), 4.94 (dd, 1H, **CH**, *J* 6.20 Hz and 2.40 Hz, H-13), 4.22-4.25 (m, 1H, **CH**, H-14), 3.50 (t, 2H, **CH**<sub>2</sub>, *J* 5.30 Hz, H-15), 3.11 (t, 1H, **CH**, *J* 2.50 Hz, H-20), 1.53 (s, 3H, **CH**<sub>3</sub>, H-17'), 1.31 (s, 3H, **CH**<sub>3</sub>, H-17). <sup>13</sup>C-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 150 MHz): 170.8 (C=O, C-21), 152.4 (Ar-C, C-4 and C-6), 151.6 (Ar-C, C-2), 144.2 (Ar-C, C-8), 131.3 (Ar-C, C-22 and C-23), 128.4 (Ar-C, C-24), 128.2 (Ar-C, C-25), 125.8 (Ar-C, C-5), 113.1 (C, C-16), 89.8 (CH, C-11), 86.8 (CH, C-14), 83.5 (CH, C-12), 81.5 (CH, C-13), 79.6 (C, C-19), 74.3 (CH, C-20), 61.4 (CH<sub>2</sub>, C-15), 37.1 (CH<sub>2</sub>, C-18), 27.2 (CH<sub>3</sub>, C-17'), 25.2 (CH<sub>3</sub>, C-17).

***N*-((*E*)-9-((3*aR*,4*R*,6*R*,6*aR*)-6-(hydroxymethyl)-2,2-dimethyltetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)-1-(prop-2-yn-1-yl)-1,9-dihydro-6*H*-purin-6-ylidene)benzamide, 70**



<sup>1</sup>H-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 600 MHz): 8.53 (s, 1H, Ar-H, H-2), 8.14 (s, 1H, Ar-H, H-8), 8.00 (dd, 2H, Ar-H, *J* 7.90 Hz and 1.30 Hz, H-23), 7.53 (t, 1H, Ar-H, *J* 7.26 Hz, H-25), 7.43 (t, 2H, Ar-H, *J* 7.80 Hz H-24), 6.05 (d, 1H, CH, *J* 2.90 Hz, H-11), 5.24 (dd, 1H, CH, *J* 6.10 Hz and 2.64 Hz, H-12), 5.07 (t, 1H, O-H, *J* 5.30 Hz), 5.00 (d, 2H, CH<sub>2</sub>, 2.40 Hz, H-18), 4.90 (dd, 1H, CH, *J* 6.20 Hz and 2.40 Hz, H-13), 4.19-4.22 (m, 1H, CH, H-14), 3.54 (t, 2H, CH<sub>2</sub>, *J* 2.40 Hz, H-20), 3.49 (t, 2H, CH<sub>2</sub>, *J* 5.30 Hz, H-15), 1.52 (s, 3H, CH<sub>3</sub>, H-17'), 1.31 (s, 3H, CH<sub>3</sub>, H-17). <sup>13</sup>C-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 150 MHz): 175.8 (C=O, C-21), 147.9 (Ar-C, C-2), 144.4 (Ar-C, C-6), 144.2 (Ar-C, C-4), 139.2 (Ar-C, C-8), 131.8 (Ar-C, C-22 and C-25), 129.1 (Ar-C, C-23), 128.3 (Ar-C, C-24), 120.9 (Ar-C, C-5), 113.1 (C, C-16), 90.0 (CH, C-11), 87.0 (CH, C-14), 84.0 (CH, C-12), 81.3 (CH, C-13), 78.9 (C, C-19), 76.0 (CH, C-20), 61.4 (CH<sub>2</sub>, C-15), 37.8 (CH<sub>2</sub>, C-18), 23.1 (CH<sub>3</sub>, C-17'), 19.3 (CH<sub>3</sub>, C-17).

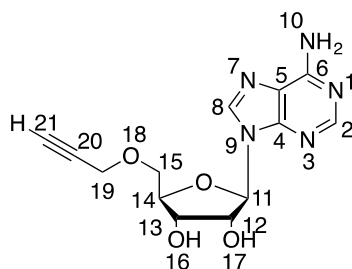
**9-((3*aR*,4*R*,6*R*,6*aR*)-2,2-dimethyl-6-((prop-2-yn-1-yloxy)methyl)tetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)-9*H*-purin-6-amine, 71** <sup>170</sup>



To a solution of ((3*aR*,4*R*,6*R*,6*aR*)-6-(6-amino-9*H*-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)methanol, **66** (3.11 g, 10.0 mmol) in THF (75 mL) at 0°C, was added NaH 60% weight (560 mg, 14.0 mmol, 1.40 eq). The reaction mixture was stirred at 0°C for 4 h before propargyl bromide 80% in THF (2.08 g, 14.0 mmol, 1.40 eq) was added. The reaction mixture was then stirred at 0°C and allowed to warm up to r.t for 20 h. The mixture was then filtered and the filtrate was concentrated *in vacuo*. Purification by flash column chromatography (0 to 2% MeOH in CH<sub>2</sub>Cl<sub>2</sub>) gave the product **71** as a light yellow oil (1.42 g, 41%). R<sub>f</sub>: 0.25

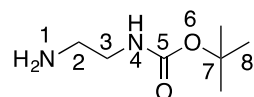
(5% MeOH in CH<sub>2</sub>Cl<sub>2</sub>). IR ( $\nu_{\max}$ , cm<sup>-1</sup>): 3280, 3130 (br, w, NH<sub>2</sub> and alkyne C-H), 2981, 2932 (w, sp<sup>3</sup>-CH), 1641, 1595, 1571 (m, adenine C=C, C=N). MS (ESI)  $m/z$  (%): 346 (100), 347 (20), 242 (90). <sup>1</sup>H-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 600 MHz): 8.30 (s, 1H, Ar-H, H-8), 8.17 (s, 1H, Ar-H, H-2), 7.35 (br, 2H, N-H), 6.16 (d, 1H, CH,  $J$  2.83 Hz, H-11), 5.41 (dd, 1H, CH,  $J$  6.00 Hz and 2.77 Hz, H-12), 4.98 (dd, 1H, CH,  $J$  6.29 Hz and 2.42 Hz, H-13), 4.31 (br, 1H, CH, H-14), 4.13-4.14 (m, 2H, CH, H-19), 3.55-3.56 (m, 1H, CH<sub>2</sub>, H-15), 3.44-3.45 (s, 1H, CH, H-21), 1.56 (s, 3H, CH<sub>3</sub>, H-17'), 1.35 (s, 3H, CH<sub>3</sub>, H-17). (Identical by <sup>1</sup>H-NMR to the literature <sup>170</sup>) <sup>13</sup>C-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 150 MHz): 155.6 (Ar-C, C-6), 153.3 (Ar-C, C-2), 149.6 (Ar-C, C-4), 139.5 (Ar-C, C-8), 120.0 (Ar-C, C-5), 114.4 (C, C-16), 91.4 (CH, C-11), 85.8 (CH, C-14), 84.9 (CH, C-12), 81.9 (CH, C-13), 69.9 (CH<sub>2</sub>, C-15), 59 (CH<sub>2</sub>, C-19), 27.2 (CH<sub>3</sub>, C-17), 25.4 (CH<sub>3</sub>, C-17'). HRMS C<sub>16</sub>H<sub>20</sub>N<sub>5</sub>O<sub>4</sub>: calc. 346.1515, found 346.1516, error: 0.30 ppm.

**(2R,3R,4S,5R)-2-(6-amino-9H-purin-9-yl)-5-((prop-2-yn-1-yloxy)methyl)tetrahydrofuran-3,4-diol, 72** <sup>171</sup>



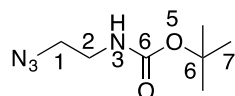
To a solution of 9-((3aR,4R,6R,6aR)-2,2-dimethyl-6-((prop-2-yn-1-yloxy)methyl)tetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)-9H-purin-6-amine, **71** (725 mg, 2.10 mmol) in *t*-BuOH (7 mL) and water (23 mL) at r.t, was added TFA (9 mL). The reaction mixture was stirred at r.t for 24 h and then concentrated *in vacuo*. The crude oil was redissolved in water and freeze-dried to afford the product as a colourless/white foam (639 mg, quantitative) which was used immediately in the next step. IR ( $\nu_{\max}$ , cm<sup>-1</sup>): 3364 (m), 3282 (s, alkyne C-H and NH<sub>2</sub>), 3120 (m, OH), 2960 (m, sp<sup>3</sup>-CH), 1623-1669 (m, adenine C=C, C=N). <sup>1</sup>H-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 600 MHz): 8.48 (s, 1H, Ar-H, H-8), 8.33 (s, 1H, Ar-H, H-2), 5.94 (d, 1H, CH,  $J$  5.18 Hz, H-11), 4.56 (dd, 1H, CH,  $J$  5.18 Hz and 5.05 Hz, H-12), 4.20-4.21 (m, 2H, CH<sub>2</sub>, H-19), 4.16 (t, 1H, CH,  $J$  4.65 Hz, H-13), 4.06-4.08 (m, 1H, CH, H-14), 3.73 (dd, 1H, CH<sub>2</sub>,  $J$  10.76 Hz and 3.72 Hz, H-15), 3.64 (dd, 1H, CH<sub>2</sub>,  $J$  10.76 Hz and 4.91 Hz, H-15), 3.48 (t, 1H, CH,  $J$  2.39 Hz, H-21). <sup>13</sup>C-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 150 MHz): 158 (Ar-C, C-6), 153 (Ar-C, C-4), 149 (Ar-C, C-2), 140.6 (Ar-C, C-8), 118.8 (Ar-C, C-5), 87.7 (CH, C-11), 83.2 (CH, C-14), 80.1 (CH, C-20), 77.6 (CH, C-21), 73.7 (CH, C-12), 70.8 (CH, C-13), 69.6 (CH<sub>2</sub>, C-15), 57.9 (CH<sub>2</sub>, C-19). MS (ESI)  $m/z$  (%): 344 (20), 306 (100). HRMS C<sub>13</sub>H<sub>16</sub>N<sub>5</sub>O<sub>4</sub>: calc. 306.1202, found 306.1200, error: 0.70 ppm.

***tert*-Butyl (2-aminoethyl)carbamate, **76****<sup>172</sup>



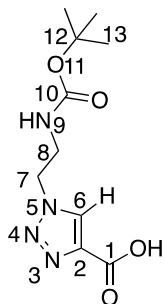
A solution of Boc<sub>2</sub>O (5.10 g, 18 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (500 mL) was added dropwise to a solution of ethylenediamine (12.0 mL, 0.18 mol, 10 eq) in CH<sub>2</sub>Cl<sub>2</sub> (120 mL) at r.t whilst stirring. The reaction mixture was continued stirring at r.t for 24 h. The white precipitate was filtered and discarded. The filtrate was concentrated *in vacuo* and saturated Na<sub>2</sub>CO<sub>3</sub> solution (200 mL) was added. After extraction with CH<sub>2</sub>Cl<sub>2</sub> (3 x 100 mL), the combined organic layers were dried (MgSO<sub>4</sub>) and concentrated *in vacuo* to afford the product as a colourless oil (2.88 g, quantitative). IR (ν<sub>max</sub>, cm<sup>-1</sup>): 3305 (m, NH), 2979, 2869 (m, sp<sup>3</sup>-CH), 1687 (s, OC=ONH), 1519 (s, OCONH). <sup>1</sup>H-NMR (CDCl<sub>3</sub>, 600 MHz): 4.93 (br, 1H, NH), 3.16-3.18 (m, br, 2H, CH<sub>2</sub>, H-3), 2.79 (t, 2H, CH<sub>2</sub>, *J* 6.0 Hz, H-2), 1.44 (s, 9H, 3 x CH<sub>3</sub>, H-8). (Identical by <sup>1</sup>H-NMR to the literature<sup>175</sup>) <sup>13</sup>C-NMR (CDCl<sub>3</sub>, 150 MHz): 156.5 (C=O, C-5), 79.4 (C, C-7), 43.4 (CH<sub>2</sub>, C-3), 42.0 (CH<sub>2</sub>, C-2), 28.5 (CH<sub>3</sub>, C-8). MS (CI) *m/z* (%): 201 (45), 161 (100), 105 (48). HRMS C<sub>7</sub>H<sub>17</sub>N<sub>2</sub>O<sub>2</sub>: calc. 161.1290, found 161.128974, error: 0.2 ppm.

***tert*-Butyl (2-azidoethyl)carbamate, **77****<sup>173</sup>



To a solution of sodium azide (6.40 g, 98.4 mmol, xs) in CH<sub>2</sub>Cl<sub>2</sub> (40 mL) and water (40 mL) at 0°C, was added (F<sub>3</sub>CSO<sub>2</sub>)<sub>2</sub>O (5.41 g, 19.2 mmol, 1.97 eq) dropwise. The reaction mixture was stirred at 0°C for 2 h then extracted with CH<sub>2</sub>Cl<sub>2</sub> (2 x 20 mL). The combined organic layers were washed once with saturated NaHCO<sub>3</sub> solution before added to a solution of *tert*-butyl (2-aminoethyl)carbamate, **76** (1.56 g, 9.74 mmol), K<sub>2</sub>CO<sub>3</sub> (2.20 g, 15.90 mmol, 1.63 eq), CuSO<sub>4</sub>·5H<sub>2</sub>O (28 mg, 0.11 mmol) in MeOH (50 mL) and water (40 mL). The total mixture was stirred at r.t for 24 h before water (50 mL) was added and pH of the solution was adjusted to 2 using HCl 37% (aq). MeOH and CH<sub>2</sub>Cl<sub>2</sub> were removed *in vacuo* followed by extraction with CH<sub>2</sub>Cl<sub>2</sub> (3 x 150 mL). The combined organic layers were washed with water (300 mL) then dried (Na<sub>2</sub>SO<sub>4</sub>) and concentrated *in vacuo* to give the product as a colourless oil (1.48 g, 81%). R<sub>f</sub>: 0.30 (CH<sub>2</sub>Cl<sub>2</sub>). IR (ν<sub>max</sub>, cm<sup>-1</sup>): 3336 (m, NH), 2978, 2933 (m, sp<sup>3</sup>-CH), 2094 (s, azide), 1686 (s, OC=ONH), 1513 (s, OCONH). <sup>1</sup>H-NMR (CDCl<sub>3</sub>, 600 MHz): 4.83 (br, 1H, N-H), 3.41 (t, 2H, CH<sub>2</sub>, *J* 5.3 Hz, H-1), 3.29-3.31 (m, 2H, CH<sub>2</sub>, H-2), 1.44 (s, 9H, 3 x CH<sub>3</sub>, H-7). (Identical by <sup>1</sup>H-NMR to the literature<sup>174</sup>) <sup>13</sup>C-NMR (CDCl<sub>3</sub>, 150 MHz): 155.8 (C=O, C-4), 79.9 (C, C-6) 51.4 (CH<sub>2</sub>, C-1), 40.2 (CH<sub>2</sub>, C-2), 38.5 (CH<sub>3</sub>, C-7). HRMS C<sub>7</sub>H<sub>15</sub>N<sub>4</sub>O<sub>2</sub>: calc. 187.11950, found 187.119168, error: 1.80 ppm.

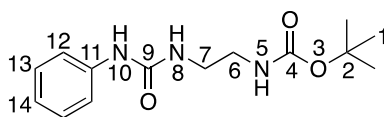
### 1-(2-((*tert*-Butoxycarbonyl)amino)ethyl)-1*H*-1,2,3-triazole-4-carboxylic acid, **78**



<sup>†</sup>This Click test reaction was not followed the general procedure but adapted from the literature procedure  
157

To a solution of propionic acid (318 mg, 4.54 mmol, 1.06 eq) in *t*-BuOH (6.0 mL), were added CuSO<sub>4</sub>·5H<sub>2</sub>O (60.0 mg, 0.24 mmol) and sodium ascorbate (200 mg, 1.02 mmol) in water (14 mL). The reaction mixture was stirred vigorously and *tert*-butyl (2-azidoethyl)carbamate, **77** (800 mg, 4.30 mmol) in *t*-BuOH (6.0 mL) followed by water (10 mL) were added to it. The reaction mixture was then stirred at r.t for 48 h before *t*-BuOH was removed *in vacuo*. As *t*-BuOH was removed *in vacuo*, a white precipitate appeared. The precipitate was filtered and washed with cold water three times and dried *in vacuo* to give the product as white flakes (826 mg, 75%). IR ( $\nu_{\max}$ , cm<sup>-1</sup>): 3381 (s, NH), 3118 (s, triazole C-H), 2979, 2932 (m, sp<sup>3</sup>-CH), 1687 (s, OC=ONH and C=OOH), 1525 (s, OCONH). <sup>1</sup>H-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 600 MHz): 13.06 (br, 1H, O-H), 8.58 (s, 1H, Ar-H, H-6), 6.99-7.01 (br, 1H, N-H), 4.43 (t, 2H, CH<sub>2</sub>, *J* 5.65 Hz, H-7), 3.37-3.40 (m, br, 2H, CH<sub>2</sub>, H-8), 1.33 (s, 9H, 3 x CH<sub>3</sub>, H-13). <sup>13</sup>C-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 150 MHz): 161.8 (C=O, C-1), 155.6 (C=O, C-10), 139 (Ar-C, C-2), 129.2 (Ar-C, C-6), 78.0 (C, C-12), 49.5 (CH<sub>2</sub>, C-7), 39.9 (CH<sub>2</sub>, C-8), 28.1 (CH<sub>3</sub>, C-13). MS (ESI) *m/z* (%): 258 (18), 257 (100). HRMS C<sub>10</sub>H<sub>17</sub>N<sub>4</sub>O<sub>4</sub>: calc. 257.1229, found 257.1250, error: 8.20 ppm.

### *tert*-Butyl (2-(3-phenylureido)ethyl)carbamate, **84** <sup>175</sup>

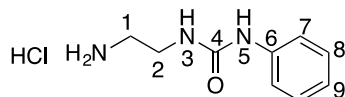


To a solution of *tert*-butyl (2-aminoethyl)carbamate, **76** (501 mg, 3.13 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (11 mL) at 0°C, was added phenyl isocyanate (372 mg, 3.13 mmol, 1 eq) in CH<sub>2</sub>Cl<sub>2</sub> (4 mL) slowly. A white precipitate formed. The mixture was stirred vigorously from 0°C to r.t for 20 h. After that, half of the solvent was removed *in vacuo*. Hexane (22 mL) was added and the precipitate was filtered and washed several times with hexane before dried *in vacuo* to afford the product as a white solid (840 mg, 97%). R<sub>f</sub>: 0.85 (EtOAc). IR ( $\nu_{\max}$ , cm<sup>-1</sup>): 3305 (m, NH), 3108 (m, triazole C-H), 2980 (w, sp<sup>3</sup>-CH), 1680 (s, C=ONH carbamate), 1642 (HNC=ONH urea), 1540 (m, Ph(C=C) and CONH). <sup>1</sup>H-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 600 MHz): 8.53 (s, br, 1H, N-H, H-10), 7.36-7.37 (m, 2H, Ar-H, H-12), 7.19-7.22 (m, 2H, Ar-H, H-13), 6.87-6.89 (m, 1H, Ar-H, H-14), 6.86-6.87 (br, 1H, N-H, H-5), 6.15 (br, 1H, N-H, H-8), 3.09-3.12 (m, 2H, CH<sub>2</sub>, H-7), 2.97-3.00



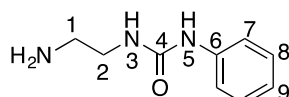
(m, 2H, CH<sub>2</sub>, H-6), 1.37 (s, 9H, 3 x CH<sub>3</sub>, H-1). <sup>13</sup>C-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 150 MHz): 155.8 (C=O, C-9), 155.3 (C=O, C-4), 140.3 (Ar-C, C-11), 128.7 (Ar-C, C-13), 121.1 (Ar-C, C-14), 117.7 (Ar-C, C-12), 40.4 (CH<sub>2</sub>, C-6), 39.1 (CH<sub>2</sub>, C-7), 28.3 (CH<sub>3</sub>, C-1). (Identical by <sup>1</sup>H- and <sup>13</sup>C-NMR to the literature <sup>160</sup>) MS (ESI) *m/z* (%): 302 (10), 281 (20), 280 (100). HRMS C<sub>14</sub>H<sub>22</sub>N<sub>3</sub>O<sub>3</sub>: calc. 280.1661, found 280.1668, error: 2.50 ppm.

### 1-(2-Aminoethyl)-3-phenylurea hydrochloride, **61** hydrochloride salt, **85.HCl** salt <sup>176</sup>



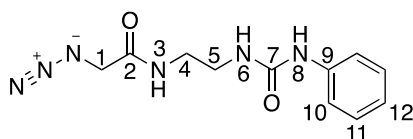
To a solution of *tert*-butyl (2-(3-phenylureido)ethyl)carbamate, **84** (251 mg, 0.96 mmol) in acetone (8.5 mL), was added HCl (6 M, 0.2 mL, 1.20 mmol, 1.25 eq). The reaction mixture was stirred at r.t for 4 h. A white precipitate started coming out of the colourless solution and the mixture was left standing at r.t overnight. The precipitate was then filtered and washed with cold EtOAc and Petroleum Ether before drying *in vacuo* to afford the product as a white solid (120 mg, 56%). IR ( $\nu_{\max}$ , cm<sup>-1</sup>): 3324, 3231 (m, NH<sub>2</sub>), 2943 (m, sp<sup>3</sup>-CH), 1656 (m, HNC=ONH), 1526-1594 (m, Ph(C=C) and CONH). <sup>1</sup>H-NMR (D<sub>2</sub>O, 600 MHz): 7.41-7.44 (m, 2H, Ar-H, H-7), 7.33-7.35 (m, 2H, Ar-H, H-8), 7.20-7.23 (m, 1H, Ar-H, H-9), 3.54 (t, 2H, CH<sub>2</sub>, *J* 5.85 Hz, H-2), 3.18 (t, 2H, CH<sub>2</sub>, *J* 5.93 Hz, H-1). (Identical by <sup>1</sup>H-NMR to the literature <sup>160</sup>) <sup>13</sup>C-NMR (D<sub>2</sub>O, 150 MHz): 159.4 (C=O, C-4), 138.4 (Ar-C, C-6), 129 (Ar-C, C-7), 125 (Ar-C, C-9), 122.3 (Ar-C, C-8), 40.6 (CH<sub>2</sub>, C-1), 38 (CH<sub>2</sub>, C-2). MS (CI) *m/z* (%): 180 (100).

### 1-(2-Aminoethyl)-3-phenylurea, **85**



To a suspension of *tert*-butyl (2-(3-phenylureido)ethyl)carbamate, **84** (161 mg, 0.61 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (7.5 mL), added TFA (1.5 mL, xs) dropwise. The reaction mixture was stirred at r.t for 4 h before left standing in fridge overnight. The solvent was then removed *in vacuo* and the remaining colourless oil was dried *in vacuo* for 3 h to remove any TFA residue before using straight in next step reaction.

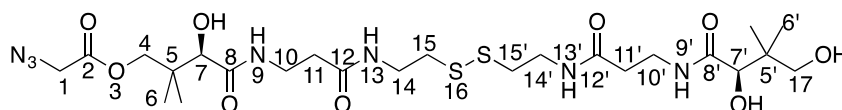
### 2-Azido-*N*-(2-(3-phenylureido)ethyl)acetamide, **86**



To a solution of 2-azidoacetic acid, **63** (50 mg, 0.50 mmol, 0.90 eq) in CH<sub>2</sub>Cl<sub>2</sub> (12.0 mL) at 0°C, HOBt (83.50 mg, 0.62 mmol, 1.12 eq) and EDC (118 mg, 0.62 mmol, 1.12 eq) were added. The reaction mixture was stirred at 0°C for 15 min. Meanwhile, 1-(2-aminoethyl)-3-

phenylurea, **85**, prepared from *tert*-butyl (2-(3-phenylureido)ethyl)carbamate, **84** and TFA, was dissolved in CH<sub>2</sub>Cl<sub>2</sub> (6.0 mL) then NEt<sub>3</sub> (0.96 mL) was added. The mixture was sonicated for 15 min to neutralise any TFA left. This solution was then added to the mixture containing 2-azidoacetic acid. The combined reaction mixture was stirred from 0°C to r.t for 20 h before concentrated *in vacuo*. Purification by flash column chromatography (20 to 50% EtOAc in Petroleum Ether then EtOAc) gave the product **86** as a colourless oil (92 mg, 64%). R<sub>f</sub>: 0.30 (EtOAc). IR (ν<sub>max</sub>, cm<sup>-1</sup>): 3380, 3325, 3269 (m, NH), 2970, 3090 (w, sp<sup>3</sup>-CH), 2107 (s, azide), 1648 (m, HNC=ONH and C=ONH amide), 1561, 1596 (m, Ph(C=C), CONH). <sup>1</sup>H-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 600 MHz): 8.54 (s, 1H, N-H, H-8), 8.21 (br, 1H, N-H, H-3), 7.37-7.39 (m, 2H, Ar-H, H-10), 7.19-7.22 (m, 2H, Ar-H, H-11), 6.87-6.89 (m, 1H, Ar-H, H-12), 6.19-6.20 (br, 1H, N-H, H-6), 3.82 (s, 2H, CH<sub>2</sub>, H-1), 3.14-3.20 (m, 4H, 2 x CH<sub>2</sub>, H-4 and H-5). <sup>13</sup>C-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 150 MHz): 167.6 (C=O, C-2), 155.4 (C=O, C-7), 140.3 (Ar-C, C-9), 128.7 (Ar-C, C-11), 121.1 (Ar-C, C-12), 117.7 (Ar-C, C-10), 50.8 (CH<sub>2</sub>, C-1), 38.8 (CH<sub>2</sub>, C-4 and C-5). MS (CI with ammonia) *m/z* (%): 280 (50), 263 (95), 252 (53), 237 (55), 235 (100). HRMS C<sub>11</sub>H<sub>15</sub>N<sub>6</sub>O<sub>2</sub>: calc. 263.12510, found 263.125090, error: 0.05 ppm.

**(3R,22R)-3,22,24-trihydroxy-2,2,23,23-tetramethyl-4,8,17,21-tetraoxo-12,13-dithia-5,9,16,20-tetraazatetracosyl 2-azidoacetate, 87**



To a solution of 2-azidoacetic acid, **63** (80 mg, 0.79 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (14.0 mL) and THF (4.0 mL) at 0°C, were added HOBt (131 mg, 0.97 mmol, 1.23 eq) and EDC (187 mg, 0.97 mmol, 1.23 eq) and NEt<sub>3</sub> (0.20 mL). The reaction mixture was stirred at 0°C for 15 min. Then D-pantethine (220 mg, 0.40 mmol, 0.5 eq) and CH<sub>2</sub>Cl<sub>2</sub> (10 mL) were added to the reaction mixture. The combined reaction mixture was stirred vigorously from 0°C to r.t for 16 h. Although LCMS showed that most of the starting material D-pantethine had not reacted, the reaction mixture was worked up. The solvent was removed and water (50 mL) and CH<sub>2</sub>Cl<sub>2</sub> (50 mL) were added. The organic layer was washed with water (2 x 30 mL) and the combined aqueous layer was reduced to 6 mL. Purification by preparative HPLC *Gradient 2*, product peak at 7.3 min gave the product **87** as a sticky yellow oil (13.3 mg). <sup>1</sup>H-NMR (D<sub>2</sub>O, 600 MHz): 4.15-4.18 (m, 3H, CH<sub>2</sub>, H-17(1) and H-1 (2)), 4.06-4.08 (m, 1H, CH<sub>2</sub>, H-17), 4.04 (s, 1H, CH, H-7'), 4.00 (s, 1H, CH, H-7), 3.50-3.58 (m, 9H, 5 x CH<sub>2</sub>, H-10 (2), H-14 (2), H-4 (1)), 3.41 (d, 1H, CH<sub>2</sub>, *J* 11.24 Hz, H-4), 2.87 (t, 4H, 2 x CH<sub>2</sub>, *J* 6.31 Hz, H-15 and H-15'), 2.53 (t, 2 x CH<sub>2</sub>, *J* 6.51 Hz, H-11 and H-11'), 1.02 and 1.00 (s, 2 x CH<sub>3</sub>, H-6'), 0.94 and 0.91 (s, 2 x CH<sub>3</sub>, H-6). <sup>13</sup>C-NMR (D<sub>2</sub>O, 150 MHz): 175.8 (C=O, C-8), 175.3 (C=O, C-8'), 174.6 (C=O, C-12 and C-12'), 171.2 (C=O, C-2), 76.4 (CH, C-7'), 75.4 (CH, C-7), 71.8 (CH<sub>2</sub>, C-17), 69 (CH<sub>2</sub>, C-4), 51.0 (CH<sub>2</sub>, C-1), 39.6 (C, C-5'), 39.2 and 38.7 (CH<sub>2</sub>, C-10 and C-14), 38.4 (C, C-5), 37.1 (CH<sub>2</sub>,

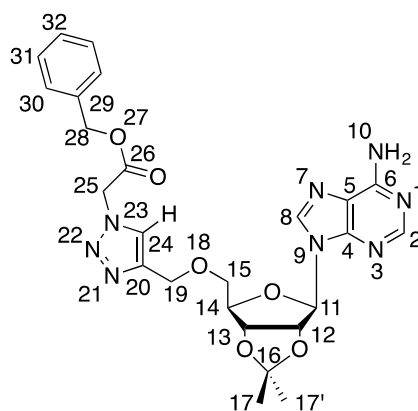
C-15), 36.0 (CH<sub>2</sub>, C-11), 21.3 and 20.2 (CH<sub>3</sub>, C-6'), 19.7 and 20.2 (CH<sub>3</sub>, C-6). HRMS C<sub>24</sub>H<sub>44</sub>N<sub>7</sub>O<sub>9</sub>S<sub>2</sub>: calc. 638.2636, found 638.2636, error: 0.11 ppm.

**General procedure for Click reaction forming 1,2,3-triazole ring system**

CuSO<sub>4</sub>·5H<sub>2</sub>O = 0.8 mol%, Sodium Ascorbate = 8 mol% from freshly made 0.05 M solutions in water.

To a solution of the alkyne in *t*-BuOH or water or both (according to solubility), were added CuSO<sub>4</sub>·5H<sub>2</sub>O 0.05 M and sodium ascorbate 0.05 M stock solutions with the ratio 0.8 mol% and 8 mol% respectively followed by a solution of the azide (1.0 eq) in *t*-BuOH or water or both (according to solubility) (not in inert conditions). The overall concentration of the azide/alkyne was 0.05 M or 0.06 M and the ratio of *t*-BuOH : H<sub>2</sub>O = 1:1 or 1:2. The reaction mixture was stirred at r.t and monitored by LCMS. In case the progress of the reaction was slow, inert condition was applied (O<sub>2</sub>-free) and ligand THPTA = 0.8 mol% was added. The reaction mixture was then concentrated *in vacuo* and purification by either flash column chromatography or preparative HPLC afforded the product.

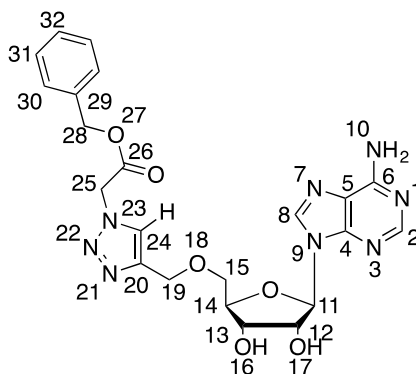
**Benzyl 2-(4-(((3*aR*,4*R*,6*R*,6*aR*)-6-(6-amino-9*H*-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)methoxy)methyl)-1*H*-1,2,3-triazol-1-yl)acetate, **82****



9-((3*aR*,4*R*,6*R*,6*aR*)-2,2-dimethyl-6-((prop-2-yn-1-yloxy)methyl)tetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)-9*H*-purin-6-amine, **71** (588 mg, 1.70 mmol), CuSO<sub>4</sub>·5H<sub>2</sub>O (0.26 mL of 0.05 M stock solution, 13 μmol, 0.8 mol%), sodium ascorbate (2.46 mL of 0.05 M stock solution, 123 μmol, 8 mol%), and benzyl 2-azidoacetate, **81** (329 mg, 1.72 mmol, 1.01 eq) were dissolved in *t*-BuOH (14 mL) and water (14 mL). After stirring at r.t for 24 h, LCMS indicated 96% completion of the reaction. The reaction mixture was then concentrated *in vacuo* and purification by flash column chromatography (2 to 3.3% MeOH in CH<sub>2</sub>Cl<sub>2</sub>) afforded the product as a colourless oil, forming foam upon drying *in vacuo* (757 mg, 83%). R<sub>f</sub>: 0.30 (2% MeOH in CH<sub>2</sub>Cl<sub>2</sub>). IR (ν<sub>max</sub>, cm<sup>-1</sup>): 3326 (m, NH<sub>2</sub>), 3144 (m, triazole C-H), 2984, 2938 (m, sp<sup>3</sup>-CH), 1750 (s, C=O ester), 1637, 1596 (s, adenine C=C, C=N). <sup>1</sup>H-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 600 MHz): 8.28 (s, 1H, Ar-H, H-8), 8.14 (s, 1H, Ar-H, H-2), 8.08 (s, 1H, Ar-H, H-24), 7.34-7.38

(m, 5H, Ar-**H**, H-30, H-31 and H-32), 7.33-7.34 (br, 2H, N-**H**), 6.14 (d, 1H, **CH**, *J* 2.65 Hz, H-11), 5.46 (s, 2H, **CH**<sub>2</sub>, H-25), 5.39 (dd, 1H, **CH**, *J* 6.20 Hz and 2.60 Hz, H-12), 5.20 (s, 2H, **CH**<sub>2</sub>, H-28), 4.95 (dd, 1H, **CH**, *J* 6.10 Hz and 2.90 Hz, H-13), 4.54 (s, 2H, **CH**<sub>2</sub>, H-19), 4.30-4.31 (m, 1H, **CH**, H-14), 3.63-3.65 (m, 1H, **CH**<sub>2</sub>, H-15), 3.55-3.57 (m, 1H, **CH**<sub>2</sub>, H-15), 1.53 (s, 3H, **CH**<sub>3</sub>, H-17), 1.31 (s, 3H, **CH**<sub>3</sub>, H-17'). <sup>13</sup>C-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 150 MHz): 167 (C=O, C-26), 156 (Ar-C, C-6), 152.8 (Ar-C, C-2), 149 (Ar-C, C-4), 143.6 (Ar-C, C-20), 139.6 (Ar-C, C-8), 135.6 (Ar-C, C-29), 128 (Ar-C, C-30, C-31, C-32), 125.5 (Ar-C, C-24), 119 (Ar-C, C-5), 113.5 (C, C-16), 89.3 (CH, C-11), 84.5 (CH, C-14), 83.4 (CH, C-12), 81.6 (CH, C-13), 69.7 (CH<sub>2</sub>, C-15), 66.8 (CH<sub>2</sub>, C-28), 63.6 (CH<sub>2</sub>, C-19), 50.3 (CH<sub>2</sub>, C-25), 27.0 (CH<sub>3</sub>, C-17), 25.2 (CH<sub>3</sub>, C-17'). MS (ESI) *m/z* (%): 538 (70), 537 (100). HRMS C<sub>25</sub>H<sub>29</sub>N<sub>8</sub>O<sub>6</sub>: calc. 537.2210, found 537.2203, error: 1.30 ppm.

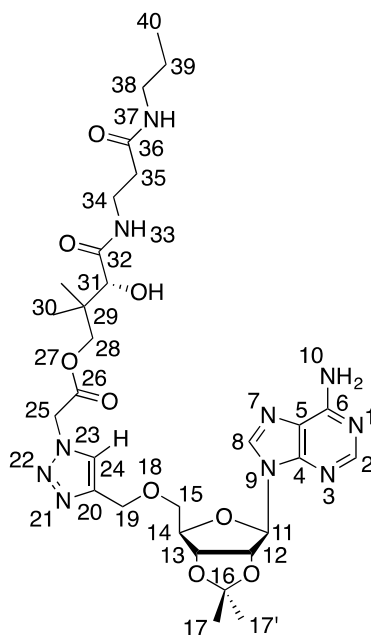
**Benzyl 2-(4-(((2*R*,3*S*,4*R*,5*R*)-5-(6-amino-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methoxy)methyl)-1*H*-1,2,3-triazol-1-yl)acetate, **14****



To a solution of benzyl 2-(4-(((3*aR*,4*R*,6*R*,6*aR*)-6-(6-amino-9*H*-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)methoxy)methyl)-1*H*-1,2,3-triazol-1-yl)acetate, **82** (910 mg, 1.70 mmol) in *t*-BuOH (30 mL) and water (30 mL), was added TFA (17 mL). The reaction mixture was stirred at r.t for 14 h. The reaction mixture was then concentrated *in vacuo* and purification by flash column chromatography (3 to 8% MeOH in CH<sub>2</sub>Cl<sub>2</sub>) afforded the product as a colourless oil, forming foam upon drying *in vacuo* (686 mg, 82%). R<sub>f</sub>: 0 (6% MeOH in CH<sub>2</sub>Cl<sub>2</sub>). IR (ν<sub>max</sub>, cm<sup>-1</sup>): 3110-3204 (br, NH<sub>2</sub>, triazole C-H), 2980, 2906 (m, sp<sup>3</sup>-CH), 1749 (s, C=O ester), 1643, 1574 (m, adenine C=C, C=N and Ph(C=C)). <sup>1</sup>H-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 600 MHz): 8.31 (s, 1H, Ar-**H**, H-8), 8.14 (s, 1H, Ar-**H**, H-2), 8.13 (s, 1H, Ar-**H**, H-24), 7.38-7.39 (m, 5H, Ar-**H**, H-30, H-31 and H-32), 5.90 (d, 1H, **CH**, *J* 5.50 Hz, H-11), 5.51 (d, 1H, O-**H**, *J* 6.30 Hz, H-16), 5.47 (s, 2H, **CH**<sub>2</sub>, H-25), 5.30 (d, 1H, O-**H**, *J* 4.90 Hz, H-17), 5.20 (s, 2H, **CH**<sub>2</sub>, H-28), 4.60 (s, 2H, **CH**<sub>2</sub>, H-19), 4.57-4.60 (m, 1H, **CH**, H-12), 4.13-4.15 (m, 1H, **CH**, H-13), 4.02-4.05 (m, 1H, **CH**, H-14), 3.71-3.74 (m, 1H, **CH**<sub>2</sub>, H-15), 3.63-3.66 (m, 1H, **CH**<sub>2</sub>, H-15). <sup>13</sup>C-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 150 MHz): 167.8 (C=O, C-26), 156 (Ar-C, C-6), 152.8 (Ar-C, C-2), 149.5 (Ar-C, C-4), 143.7 (Ar-C, C-20), 139.4 (Ar-C, C-8), 135.5 (Ar-C, C-29), 128 (Ar-C, C-30, C-31, C-32), 125.4 (Ar-C, C-24), 119 (Ar-C, C-5), 87.2 (CH, C-11), 83.1 (CH, C-14), 73.3

(CH, C-12), 70.6 (CH, C-13), 70.1 (CH<sub>2</sub>, C-15), 66.8 (CH<sub>2</sub>, C-28), 63.7 (CH<sub>2</sub>, C-19), 50.3 (CH<sub>2</sub>, C-25). MS (ESI) *m/z* (%): 498 (25), 497 (100). HRMS C<sub>22</sub>H<sub>25</sub>N<sub>8</sub>O<sub>6</sub>: calc. 497.1897, found 497.1911, error: 2.80 ppm.

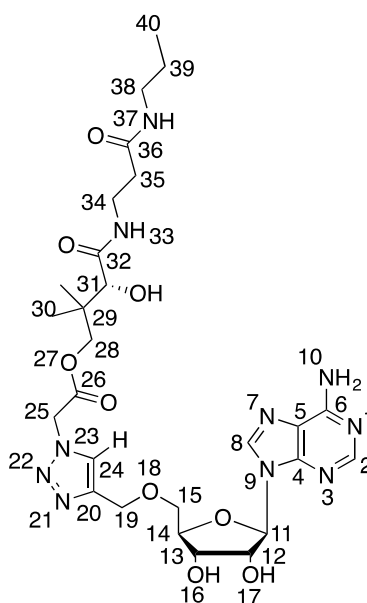
**(R)-3-hydroxy-2,2-dimethyl-4-oxo-4-((3-oxo-3-(propylamino)propyl)amino)butyl 2-(((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)methoxy)methyl)-1H-1,2,3-triazol-1-yl)acetate, 17**



9-((3aR,4R,6R,6aR)-2,2-dimethyl-6-((prop-2-yn-1-yloxy)methyl)tetrahydrofuro[3,4-d][1,3]dioxol-4-yl)-9H-purin-6-amine, **71** (53 mg, 0.15 mmol), CuSO<sub>4</sub>·5H<sub>2</sub>O (24 μL of 0.05 M stock solution, 1.20 μmol, 0.8 mol%), sodium ascorbate (0.24 mL of 0.05 M stock solution, 12 μmol, 8 mol%), and (R)-3-hydroxy-2,2-dimethyl-4-oxo-4-((3-oxo-3-(propylamino)propyl)amino)butyl 2-azidoacetate, **64** (64 mg, 0.15 mmol, 1.00 eq) were dissolved in *t*-BuOH (1.5 mL) and water (1.5 mL). After stirring at r.t for 15 h, LCMS indicated 60% completion of the reaction. The reaction mixture was then concentrated *in vacuo* and purification by flash column chromatography (3 to 8% MeOH in CH<sub>2</sub>Cl<sub>2</sub>) afforded the product as a colourless oil (58 mg, 56%). R<sub>f</sub>: 0.10 (10% MeOH in CH<sub>2</sub>Cl<sub>2</sub>). IR (ν<sub>max</sub>, cm<sup>-1</sup>): 3360, 3326 (br, m, triazole C-H and NH<sub>2</sub>), 2942 (w, sp<sup>3</sup>-CH), 1754 (s, C=O ester), 1703, 1687 (s, C=ONH), 1619, 1572 (m, Ph(C=C) and adenine, C=C, C=N). <sup>1</sup>H-NMR (CD<sub>3</sub>OD, 600 MHz): 8.25 (s, 1H, Ar-H, H-8), 8.18 (s, 1H, Ar-H, H-2), 7.90 (s, 1H, Ar-H, H-24), 6.18 (d, 1H, CH, *J* 2.57 Hz, H-11), 5.36 (dd, 1H, CH, *J* 6.05 Hz and 2.50 Hz, H-12), 5.34 (s, 2H, CH<sub>2</sub>, H-25), 5.02 (dd, 1H, CH, *J* 6.00 Hz and 2.22 Hz, H-13), 4.59 (s, 2H, CH<sub>2</sub>, H-19), 4.47-4.48 (m, 1H, CH, H-14), 4.03 and 4.12 (d, 2H, CH<sub>2</sub>, *J* 10.71 Hz, H-28), 3.78 (s, 1H, CH, H-31), 3.72-3.74 (m, 1H, CH<sub>2</sub>, H-15), 3.66-3.68 (m, 1H, CH<sub>2</sub>, H-15), 3.41-3.51 (m, 2H, CH<sub>2</sub>, H-34), 3.08-3.13 (m, 2H, CH<sub>2</sub>, H-38), 2.41 (t, 2H, CH<sub>2</sub>, *J* 6.77 Hz, H-35), 1.59 (s, 3H, CH<sub>3</sub>, H-17), 1.47-1.53 (m, 2H, CH<sub>2</sub>, H-39), 1.38 (s, 3H, CH<sub>3</sub>, H-17'), 0.95 (s, 3H, CH<sub>3</sub>, H-30), 0.91 (s, 3H, CH<sub>3</sub>, H-30), and

0.89 (t, 3H, CH<sub>3</sub>, *J* 7.00 Hz, H-40). <sup>13</sup>C-NMR (CD<sub>3</sub>OD, 150 MHz): 175.2 (C=O, C-32), 173.6 (C=O, C-36), 168 (C=O, C-26), 157.1 (Ar-C, C-6), 153.6 (Ar-C, C-2), 150.5 (Ar-C, C-4), 145.4 (Ar-C, C-20), 141.4 (Ar-C, C-8), 126.6 (Ar-C, C-24), 120.3 (Ar-C, C-5), 115 (C, C-16), 92.5 (CH, C-11), 87.3 (CH, C-14), 85.9 (CH, C-12), 83.4 (CH, C-13), 75.7 (CH, C-31), 72.7 (CH<sub>2</sub>, C-28), 71.5 (CH<sub>2</sub>, C-15), 64.9 (CH<sub>2</sub>, C-19), 51.8 (CH<sub>2</sub>, C-25), 42.2 (CH<sub>2</sub>, C-38), 39.3 (C, C-29), 36.4 (CH<sub>2</sub>, C-34), 36.36 (CH<sub>2</sub>, C-35), 27.5 (CH<sub>3</sub>, C-17), 25.4 (CH<sub>3</sub>, C-17'), 23.6 (CH<sub>2</sub>, C-39), 21.7 (CH<sub>3</sub>, C-30), 25.5 (CH<sub>3</sub>, C-30), 11.7 (CH<sub>3</sub>, C-40). MS (ESI) *m/z* (%): 711 (20), 691 (10), 690 (40), 689 (100). HRMS C<sub>30</sub>H<sub>45</sub>N<sub>10</sub>O<sub>9</sub>: calc. 689.3371, found 689.3378, error: 1.01 ppm.

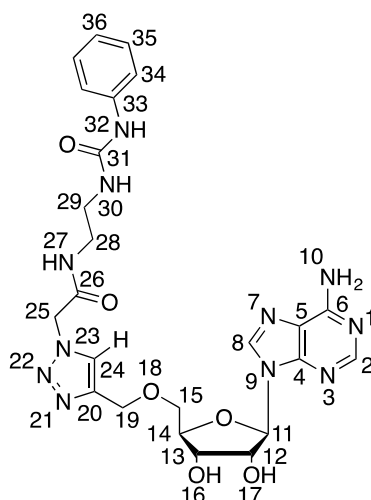
**(*R*)-3-hydroxy-2,2-dimethyl-4-oxo-4-((3-oxo-3-(propylamino)propyl)amino)butyl 2-(4-(((2*R*,3*S*,4*R*,5*R*)-5-(6-amino-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methoxy)methyl)-1*H*-1,2,3-triazol-1-yl)acetate, **16****



(2*R*,3*R*,4*S*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((prop-2-yn-1-yloxy)methyl)tetrahydrofuran-3,4-diol, **72** (26 mg, 86 μmol), CuSO<sub>4</sub>·5H<sub>2</sub>O (15 μL of 0.05 M stock solution, 0.75 μmol, 0.8 mol%), sodium ascorbate (0.15 mL of 0.05 M stock solution, 7.5 μmol, 8 mol%) and (*R*)-3-hydroxy-2,2-dimethyl-4-oxo-4-((3-oxo-3-(propylamino)propyl)amino)butyl 2-azidoacetate, **64** (29 mg, 85 μmol, 0.99 eq) were dissolved in *t*-BuOH (0.8 mL) and water (0.8 mL) in Argon condition (O<sub>2</sub>-free). THPTA ligand (1.0 mg, 2.30 μmol) was added. After stirring at r.t for 2 h, the reaction mixture was then concentrated *in vacuo* and purification by preparative HPLC, *Gradient 3*, product peak at 11.6 min, afforded the product as a colourless oil (49 mg, 90%). IR ( $\nu_{\max}$ , cm<sup>-1</sup>): 3305 (m, NH<sub>2</sub>), 3108 (m, triazole C-H), 2965 (m, sp<sup>3</sup>-CH), 1748 (s, C=O ester), 1643 (m, adenine C=C, C=N and C=ONH), 1535 (m, CONH). <sup>1</sup>H-NMR (D<sub>2</sub>O, 600 MHz): 8.49 (s, 1H, Ar-H, H-8), 8.44 (s, 1H, Ar-H, H-2), 8.14 (s, 1H, Ar-H, H-24), 6.18 (d, 1H, CH, *J* 4.68 Hz, H-11), 5.47 (s, 2H, CH<sub>2</sub>, H-25), 4.81-4.97 (m, 1H, CH, H-12), 4.76 (dd, 2H, CH<sub>2</sub>, *J* 12.23 Hz and 2.91 Hz, H-19), 4.50 (t, 1H, CH, *J* 4.92 Hz, H-13), 4.37-4.39 (m, 1H, CH, H-14),

4.03 and 4.12 (d, 2H,  $\text{CH}_2$ ,  $J$  10.49 Hz, H-28), 3.94 (dd, 1H,  $\text{CH}_2$ ,  $J$  11.14 Hz and 2.62 Hz, H-15), 3.88 (dd, 1H,  $\text{CH}_2$ ,  $J$  11.14 Hz and 4.26 Hz, H-15), 3.83 (s, 1H,  $\text{CH}$ , H-31), 3.45-3.53 (m, 2H,  $\text{CH}_2$ , H-34), 3.12 (dt, 2H,  $\text{CH}_2$ ,  $J$  6.81 Hz and 2.06 Hz, H-38), 2.49 (t, 2H,  $\text{CH}_2$ ,  $J$  6.40 Hz, H-35), 1.46-1.52 (m, 2H,  $\text{CH}_2$ , H-39), 0.95 (s, 3H,  $\text{CH}_3$ , H-30), 0.91 (s, 3H,  $\text{CH}_3$ , H-30), and 0.87 (t, 3H,  $\text{CH}_3$ ,  $J$  7.00 Hz, H-40).  $^{13}\text{C}$ -NMR ( $\text{D}_2\text{O}$ , 150 MHz): 175.0 (C=O, C-32), 174.3 (C=O, C-36), 168.9 (C=O, C-26), 150.7 (Ar-C, C-6), 149 (Ar-C, C-4), 145.4 (Ar-C, C-2), 144.7 (Ar-C, C-20), 143.2 (Ar-C, C-8), 127.1 (Ar-C, C-24), 119 (Ar-C, C-5), 89.2 (CH, C-11), 84.6 (CH, C-14), 75.3 (CH, C-12), 75.0 (CH, C-31), 72.2 ( $\text{CH}_2$ , C-28), 71.1 (CH, C-13), 69.8 ( $\text{CH}_2$ , C-15), 63.9 ( $\text{CH}_2$ , C-19), 51.6 ( $\text{CH}_2$ , C-25), 41.9 ( $\text{CH}_2$ , C-38), 36.1 ( $\text{CH}_2$ , C-35), 38.2 (C, C-29), 35.9 ( $\text{CH}_2$ , C-34), 22.4 ( $\text{CH}_2$ , C-39), 22.3 ( $\text{CH}_3$ , C-30), 20.7 ( $\text{CH}_3$ , C-30), 11.3 ( $\text{CH}_3$ , C-40). MS (ESI)  $m/z$  (%): 671 (20), 650 (30), 649 (100). HRMS  $\text{C}_{27}\text{H}_{41}\text{N}_{10}\text{O}_9$ : calc. 649.3058, found 649.3055, error: 0.46 ppm.

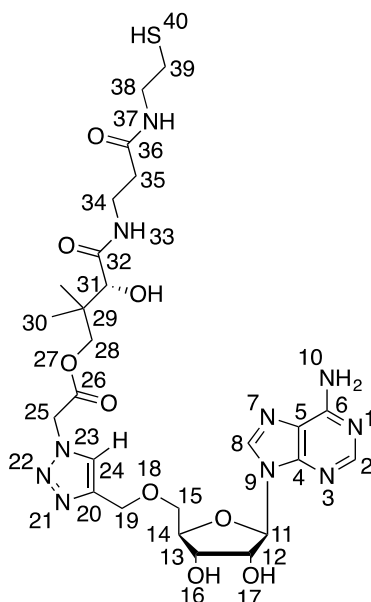
**2-(4-(((2*R*,3*S*,4*R*,5*R*)-5-(6-amino-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methoxy)methyl)-1*H*-1,2,3-triazol-1-yl)-*N*-(2-(3-phenylureido)ethyl)acetamide, **15****



(2*R*,3*R*,4*S*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((prop-2-yn-1-yloxy)methyl)tetrahydrofuran-3,4-diol, **72** (56.5 mg, 0.19 mmol),  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$  (30  $\mu\text{L}$  of 0.05 M stock solution, 1.5  $\mu\text{mol}$ , 0.8 mol%), sodium ascorbate (0.30 mL of 0.05 M stock solution, 15  $\mu\text{mol}$ , 8 mol%) and 2-azido-*N*-(2-(3-phenylureido)ethyl)acetamide, **86** (55.5 mg, 0.20 mmol, 1.05 eq) were dissolved in *t*-BuOH (1.71 mL) and water (1.71 mL) in Argon condition ( $\text{O}_2$ -free). THPTA ligand (1.0 mg, 2.30  $\mu\text{mol}$ ) was added. After stirring at r.t for 12 h, the reaction mixture was then concentrated *in vacuo* and purification by preparative HPLC *Gradient I*, product peak at 3.6 min, afforded the product as a colourless oil, forming foam upon drying *in vacuo* (43 mg, 40%). IR ( $\nu_{\text{max}}$ ,  $\text{cm}^{-1}$ ): 3279 (m,  $\text{NH}_2$ ), 3097 (m, triazole C-H), 2965 (m,  $\text{sp}^3\text{-CH}$ ), 1659 (m, adenine C=N and CONH), 1596 (m, adenine C=C, and CONH).  $^1\text{H}$ -NMR ( $\text{D}_2\text{O}$ , 600 MHz): 8.37 (s, 1H, Ar-H, H-8), 8.33 (s, 1H, Ar-H, H-2), 8.01 (s, 1H, Ar-H, H-24), 7.27-7.20 (m, 2H, Ar-H, H-35), 7.16-7.17 (m, 2H, Ar-H, H-34), 7.06-7.09 (m, 1H, Ar-H, H-36), 6.11 (d, 1H,  $\text{CH}$ ,  $J$  5.06 Hz, H-11), 5.28 (dd, 2H,  $\text{CH}_2$ ,  $J$  16.50 Hz and 1.80 Hz, H-25), 4.77 (d, 1H,  $\text{CH}$ ,  $J$  5.80 Hz, H-12), 4.65 (dd,

2H, CH<sub>2</sub>, *J* 14.40 Hz and 1.50 Hz, and *J* 14.40 Hz and 2.30 Hz, H-19), 4.46 (t, 1H, CH, *J* 5.0 Hz, H-13), 4.34-4.35 (m, 1H, CH, H-14), 3.86 (dd, 2H, CH<sub>2</sub>, *J* 11.30 Hz and 4.40 Hz, and *J* 11.30 Hz and 2.70 Hz, H-15), 3.41-3.43 (m, 2H, CH<sub>2</sub>, H-28), 3.34-3.35 (m, 2H, CH<sub>2</sub>, H-29). <sup>13</sup>C-NMR (D<sub>2</sub>O, 150 MHz): 168 (C=O, C-26), 158 (C=O, C-31), 151 (Ar-C, C-6), 146.5 (Ar-C, C-2 and C-20), 142.6 (Ar-C, C-8), 138.5 (Ar-C, C-33), 129.9 (Ar-C, C-35 and C-36), 124.6 (Ar-C, C-24), 121.7 (Ar-C, C-34), 121 (Ar-C, C-5), 88.9 (CH, C-11), 84.5 (CH, C-14), 74.8 (CH, C-12), 71.0 (CH, C-13), 69.8 (CH<sub>2</sub>, C-15), 63.9 (CH<sub>2</sub>, C-19), 52.9 (CH<sub>2</sub>, C-25), 40.3 (CH<sub>2</sub>, C-28), 39.7 (CH<sub>2</sub>, C-29). MS (ESI) *m/z* (%): 606 (10), 569 (30), 568 (100). HRMS C<sub>24</sub>H<sub>30</sub>N<sub>11</sub>O<sub>6</sub>: calc. 568.2394, found 568.2381, error: 2.30 ppm.

**(*R*)-3-hydroxy-4-((3-((2-mercaptoethyl)amino)-3-oxopropyl)amino)-2,2-dimethyl-4-oxobutyl 2-(4-(((2*R*,3*S*,4*R*,5*R*)-5-(6-amino-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methoxy)methyl)-1*H*-1,2,3-triazol-1-yl)acetate, **18****



\* Two Step

*Step 1*

(2*R*,3*R*,4*S*,5*R*)-2-(6-amino-9*H*-purin-9-yl)-5-((prop-2-yn-1-yloxy)methyl)tetrahydrofuran-3,4-diol, **72** (6.30 mg, 21 μmol), CuSO<sub>4</sub>·5H<sub>2</sub>O (4 μL of 0.05 M stock solution, 0.2 μmol, 0.8 mol%), sodium ascorbate (45 μL of 0.05 M stock solution, 2.2 μmol, 8 mol%) and (3*R*,22*R*)-3,22,24-trihydroxy-2,2,23,23-tetramethyl-4,8,17,21-tetraoxo-12,13-dithia-5,9,16,20-tetraazatetracosyl 2-azidoacetate, **87** (13.2 mg, 21 μmol, 1.00 eq) were dissolved in *t*-BuOH (120 μL) and water (240 μL) in Argon condition (O<sub>2</sub>-free). THPTA ligand (0.30 mg, 0.69 μmol) was added. After stirring at r.t for 12 h, LCMS indicated 50% completion of the reaction. The reaction mixture was then concentrated *in vacuo* and purification by preparative HPLC *Gradient 9*, product peak at 5 min, afforded the crude product as a light brown oil (≈18 mg, 93%). NMR showed that the ratio of starting material : product = 1:1.35 - 1:1.5.

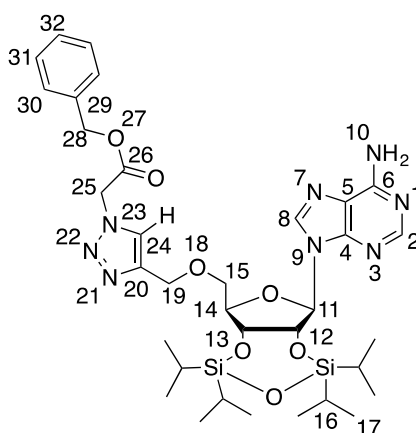


## Step 2

To the crude material from Step 1 in acetonitrile (1.20 mL) and water (1.80 mL), was added TCEP (26.5 mg, 92  $\mu$ mol, 4.50 eq). The reaction mixture was stirred at r.t for 2 h before freeze-drying to remove the solvent. Purification by preparative HPLC *Gradient 4*, product peak at 10.6 min, afforded the product as a colourless oil, forming foam upon drying *in vacuo* (13.4 mg, 95%).

$^1\text{H-NMR}$  ( $\text{D}_2\text{O}$ , 600 MHz): 8.46 (s, 1H, Ar-H, H-8), 8.40 (s, 1H, Ar-H, H-2), 8.13 (s, 1H, Ar-H, H-24), 6.17 (d, 1H, CH,  $J$  4.70 Hz, H-11), 5.47 (s, 2H,  $\text{CH}_2$ , H-25), 4.79 (1H, CH, H-12, hidden under solvent peak), 4.75-4.76 (m, 2H,  $\text{CH}_2$ , H-19), 4.50 (t, 1H, CH,  $J$  4.82 Hz, H-13), 4.37-4.39 (m, 1H, CH, H-14), 4.11 and 4.03 (d, 2H,  $\text{CH}_2$ ,  $J$  10.76 Hz, H-28), 3.94 (dd, 1H,  $\text{CH}_2$ ,  $J$  11.28 Hz and 2.77 Hz, H-15), 3.87 (dd, 1H,  $\text{CH}_2$ ,  $J$  11.28 Hz and 4.31 Hz, H-15), 3.83 (s, 1H, CH, H-31), 3.46-3.53 (m, 2H,  $\text{CH}_2$ , H-34), 3.36 (dt, 2H,  $\text{CH}_2$ ,  $J$  6.66 Hz and 2.15 Hz, H-38), 2.64 (t, 2H,  $\text{CH}_2$ ,  $J$  6.56 Hz, H-35), 2.51 (t, 2H,  $\text{CH}_2$ ,  $J$  6.46 Hz, H-39), 0.95 (s, 3H,  $\text{CH}_3$ , H-30), 0.91 (s, 3H,  $\text{CH}_3$ , H-30).  $^{13}\text{C-NMR}$  ( $\text{D}_2\text{O}$ , 150 MHz): 175.0 (C=O, C-32), 174.6 (C=O, C-36), 168.8 (C=O, C-26), 152 (Ar-C, C-6), 149.2 (Ar-C, C-4), 146.9 (Ar-C, C-2), 144.7 (Ar-C, C-20), 142.7 (Ar-C, C-8), 127.1 (Ar-C, C-24), 119.3 (Ar-C, C-5), 89.0 (CH, C-11), 84.6 (CH, C-14), 75.3 (CH, C-12), 74.9 (CH, C-31), 72.2 ( $\text{CH}_2$ , C-28), 71.1 (CH, C-13), 69.8 ( $\text{CH}_2$ , C-15), 63.9 ( $\text{CH}_2$ , C-19), 51.7 ( $\text{CH}_2$ , C-25), 42.9 ( $\text{CH}_2$ , C-38), 38.2 (C, C-29), 36.1 ( $\text{CH}_2$ , C-34), 35.9 ( $\text{CH}_2$ , C-39), 23.8 ( $\text{CH}_2$ , C-35), 20.8 ( $\text{CH}_3$ , C-30), 20.3 ( $\text{CH}_3$ , C-30). HRMS  $\text{C}_{26}\text{H}_{38}\text{N}_{10}\text{O}_9\text{S}$ : calc. 667.2617, found 667.2617, error: 0 ppm.

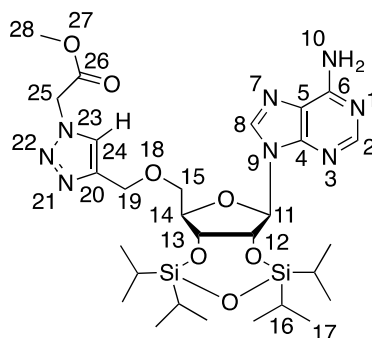
## Benzyl 2-(4-(((5*aR*,6*R*,8*R*,8*aR*)-8-(6-amino-9*H*-purin-9-yl)-2,2,4,4-tetraisopropyltetrahydrofuro[3,4-*f*][1,3,5,2,4]trioxadisilepin-6-yl)methoxy)methyl)-1*H*-1,2,3-triazol-1-yl)acetate, **89**



Benzyl 2-(4-(((2*R*,3*S*,4*R*,5*R*)-5-(6-amino-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methoxy)methyl)-1*H*-1,2,3-triazol-1-yl)acetate, **14** (353 mg, 0.71 mmol) was dissolved in pyridine (33 mL) and TIPDSCl<sub>2</sub> (0.28 mL, 0.85 mmol, 1.20 eq) was added. The reaction mixture was stirred at r.t for 18 h before the solvent was removed *in vacuo*. Purification by flash

column chromatography (0 to 2.5% MeOH in CH<sub>2</sub>Cl<sub>2</sub>) afforded the product as a colourless oil, forming foam upon drying *in vacuo* (513 mg, 98%). R<sub>f</sub>: 0.75 (6% MeOH in CH<sub>2</sub>Cl<sub>2</sub>). IR (ν<sub>max</sub>, cm<sup>-1</sup>): 3152 (m, triazole C-H), 2945, 2866 (m, sp<sup>3</sup>-CH), 1751 (s, C=O ester), 1573-1681 (m, adenine C=C, C=N). <sup>1</sup>H-NMR (CD<sub>3</sub>OD, 600 MHz): 8.33 (s, 1H, Ar-H, H-8), 8.17 (s, 1H, Ar-H, H-2), 8.01 (s, 1H, Ar-H, H-24), 7.27-7.33 (m, 5H, Ar-H, H-30, H-31 and H-32), 6.08 (d, 1H, CH, *J* 5.13 Hz, H-11), 5.38 (s, 2H, CH<sub>2</sub>, H-25), 5.21 (s, 2H, CH<sub>2</sub>, H-28), 5.01 (dd, 1H, CH, *J* 5.13 Hz and 5.01 Hz (roofing effect), H-12), 4.71 (dd, 2H, CH<sub>2</sub>, *J* 12.15 Hz and 3.38 Hz, H-19), 4.69 (dd, 1H, CH, *J* 4.63 Hz and 4.38 Hz (roofing effect), H-13), 4.25-4.27 (m, 1H, CH, H-14), 3.85 (dd, 1H, CH<sub>2</sub>, *J* 10.90 Hz and 3.26 Hz, H-15), 3.76 (dd, 1H, CH<sub>2</sub>, *J* 10.90 Hz and 3.63 Hz, H-15), 0.95-1.10 (m, 28H, 4 x CH and 8 x CH<sub>3</sub>, H-16 and H-17). <sup>13</sup>C-NMR (CD<sub>3</sub>OD, 150 MHz): 168.1 (C=O, C-26), 157.2 (Ar-C, C-6), 153.9 (Ar-C, C-2), 150.8 (Ar-C, C-4), 145.7 (Ar-C, C-20), 140.9 (Ar-C, C-8), 136.6 (Ar-C, C-29), 129.3-129.6 (Ar-C, C-30, C-31, C-32), 126.6 (Ar-C, C-24), 120.3 (Ar-C, C-5), 89.9 (CH, C-11), 85.8 (CH, C-14), 78.1 (CH, C-12), 74.4 (CH, C-13), 70.8 (CH<sub>2</sub>, C-15), 68.8 (CH<sub>2</sub>, C-28), 65.2 (CH<sub>2</sub>, C-19), 51.7 (CH<sub>2</sub>, C-25), 17.6 (CH<sub>3</sub>, C-17), 14.3 (CH, C-16). MS (ESI) *m/z* (%): 741 (15), 740 (35), 739 (100). HRMS C<sub>34</sub>H<sub>51</sub>N<sub>8</sub>O<sub>7</sub>Si<sub>2</sub>: calc. 739.3419, found 739.3430, error: 1.50 ppm.

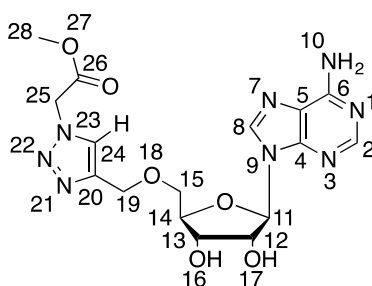
**Methyl 2-(4-(((5a*R*,6*R*,8*R*,8a*R*)-8-(6-amino-9*H*-purin-9-yl)-2,2,4,4-tetraisopropyltetrahydrofuro[3,4-*f*][1,3,5,2,4]trioxadisilepin-6-yl)methoxy)methyl)-1*H*-1,2,3-triazol-1-yl)acetate, **90****



To a solution of benzyl 2-(4-(((5a*R*,6*R*,8*R*,8a*R*)-8-(6-amino-9*H*-purin-9-yl)-2,2,4,4-tetraisopropyltetrahydrofuro[3,4-*f*][1,3,5,2,4]trioxadisilepin-6-yl)methoxy)methyl)-1*H*-1,2,3-triazol-1-yl)acetate, **89** (172 mg, 0.23 mmol) in MeOH (9 mL) and CH<sub>2</sub>Cl<sub>2</sub> (9 mL), was added Pd/C 10% (17.7 mg, 16.6 μmol). The reaction mixture was stirred at r.t for 24 h under H<sub>2</sub> condition. The mixture was then filtered and the filtrate was concentrated *in vacuo* to afford the product as a white solid (150 mg, 97%). IR (ν<sub>max</sub>, cm<sup>-1</sup>): 3283 (w, NH<sub>2</sub>), 3147 (m, triazole C-H), 2946, 2865 (m, sp<sup>3</sup>-CH), 1747 (s, C=O ester), 1600-1679 (m, adenine C=C, C=N). <sup>1</sup>H-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 600 MHz): 8.36 (s, 1H, Ar-H, H-8), 8.13 (s, 1H, Ar-H, H-2), 8.07 (s, 1H, Ar-H, H-24), 7.34 (br, 2H, N-H), 5.98 (d, 1H, CH, *J* 5.47 Hz, H-11), 5.40 (s, 2H, CH<sub>2</sub>, H-25), 5.10 (t, 1H, CH, *J* 5.20 Hz, H-12), 4.61-4.64 (m, 3H, CH and CH<sub>2</sub>, H-13 and H-19), 4.13-4.14 (m,

1H, **CH**, H-14), 3.77 (dd, 1H, **CH<sub>2</sub>**, *J* 10.86 Hz and 3.96 Hz, H-15), 3.67-3.70 (m, 4H, **CH<sub>2</sub>** (1H) and **CH<sub>3</sub>**, H-15 and H-28), 1.00-1.10 (m, 28H, 4 x **CH** and 8 x **CH<sub>3</sub>**, H-16 and H-17). <sup>13</sup>C-NMR ((CD<sub>3</sub>)<sub>2</sub>SO, 150 MHz): 168.7 (C=O, C-26), 157.3 (Ar-C, C-6), 154.0 (Ar-C, C-2), 150.7 (Ar-C, C-4), 145.7 (Ar-C, C-20), 141.0 (Ar-C, C-8), 126.6 (Ar-C, C-24), 120.2 (Ar-C, C-5), 89.7 (CH, C-11), 85.8 (CH, C-14), 78.0 (CH, C-12), 74.3 (CH, C-13), 70.7 (CH<sub>2</sub>, C-15), 65.1 (CH<sub>2</sub>, C-19), 53.3 (CH<sub>3</sub>, C-28), 51.6 (CH<sub>2</sub>, C-25), 17.4-17.9 (CH<sub>3</sub>, C-17), 14.4 (CH, C-16). MS (ESI) *m/z* (%): 685 (20), 665 (13), 664 (40), 663 (100) (15), 740 (35), 739 (100). HRMS C<sub>28</sub>H<sub>47</sub>N<sub>8</sub>O<sub>7</sub>Si<sub>2</sub>: calc. 664.3185, found 664.3180, error: 0.75 ppm.

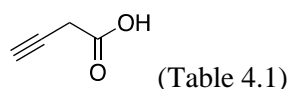
**Methyl 2-(4-(((2*R*,3*S*,4*R*,5*R*)-5-(6-amino-9*H*-purin-9-yl)-3,4-dihydroxytetrahydrofuran-2-yl)methoxy)methyl)-1*H*-1,2,3-triazol-1-yl)acetate, **91****



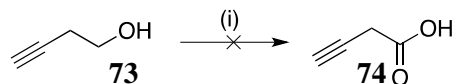
To a solution of methyl 2-(4-(((5*aR*,6*R*,8*R*,8*aR*)-8-(6-amino-9*H*-purin-9-yl)-2,2,4,4-tetraisopropyltetrahydrofuro[3,4-*f*][1,3,5,2,4]trioxadisilepin-6-yl)methoxy)methyl)-1*H*-1,2,3-triazol-1-yl)acetate, **90** (48 mg, 72 μmol) in MeOH (2 mL), was added NH<sub>4</sub>F (23 mg, 0.62 mmol, 8.60 eq). The reaction mixture was stirred at r.t for 2 h then concentrated *in vacuo*. Purification by flash column chromatography (3 to 10% MeOH in CH<sub>2</sub>Cl<sub>2</sub>) afforded the product as a colourless oil (26 mg, 86%). R<sub>f</sub>: 0.12 (10% MeOH in CH<sub>2</sub>Cl<sub>2</sub>). <sup>1</sup>H-NMR (CD<sub>3</sub>OD, 600 MHz): 8.31 (s, 1H, Ar-**H**, H-8), 8.18 (s, 1H, Ar-**H**, H-2), 8.02 (s, 1H, Ar-**H**, H-24), 7.34 (br, 2H, N-**H**), 6.06 (d, 1H, **CH**, *J* 4.80 Hz, H-11), 5.35 (s, 2H, **CH<sub>2</sub>**, H-25), 4.69-4.73 (m, 2H, **CH<sub>2</sub>**, H-19), 4.59 (t, 1H, **CH**, *J* 4.80 Hz, H-12), 4.34 (t, 1H, **CH**, *J* 4.7 Hz, H-13), 4.20-4.22 (m, 1H, **CH**, H-14), 3.86 (dd, 1H, **CH<sub>2</sub>**, *J* 10.80 Hz and 3.60 Hz, H-15), 3.78 (dd, 1H, **CH<sub>2</sub>**, *J* 10.80 Hz and 3.20 Hz, H-15), 3.77 (s, 3H, **CH<sub>3</sub>**, H-28). <sup>13</sup>C-NMR (CD<sub>3</sub>OD, 150 MHz): 168.8 (C=O, C-26), 157.0 (Ar-C, C-6), 153.9 (Ar-C, C-2), 150.3 (Ar-C, C-4), 145.8 (Ar-C, C-20), 140.9 (Ar-C, C-8), 126.6 (Ar-C, C-24), 120.3 (Ar-C, C-5), 89.7 (CH, C-11), 85.3 (CH, C-14), 76.2 (CH, C-12), 72.1 (CH, C-13), 71.0 (CH<sub>2</sub>, C-15), 65.1 (CH<sub>2</sub>, C-19), 53.3 (CH<sub>3</sub>, C-28), 51.6 (CH<sub>2</sub>, C-25). MS (ESI) *m/z* (%): 741 (15), 740 (35), 739 (100). HRMS C<sub>16</sub>H<sub>21</sub>N<sub>8</sub>O<sub>6</sub>: calc. 421.1584, found 421.1588, error: 0.95 ppm.

## Failed synthetic attempts

### But-3-ynoic acid, **74**



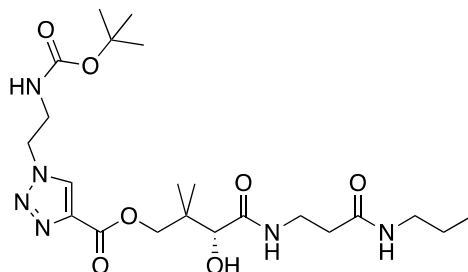
**Table 4.1**



Conditions	Results
NaIO <sub>4</sub> , K <sub>2</sub> Cr <sub>2</sub> O <sub>7</sub> , HNO <sub>3</sub> , water, r.t. <sup>150</sup>	polymerised

To water (35 mL) at 0°C, were added NaIO<sub>4</sub> (11.80 g, 55.00 mol, 2.20 eq), K<sub>2</sub>Cr<sub>2</sub>O<sub>4</sub> (0.08g, 0.25 mol, 1 eq) and HNO<sub>3</sub> (0.08 mL, 1.25 mmol). But-3-yn-1-ol, **73** (1.81g, 25.0 mol) was added slowly to the reaction mixture whilst stirring. The reaction was exothermic. The reaction mixture was stirred at 10°C for 30 min. The aqueous reaction mixture was extracted with Et<sub>2</sub>O (3 x 100 mL). The combined extract was dried (MgSO<sub>4</sub>). The solvent was removed *in vacuo* to give a brown paste.

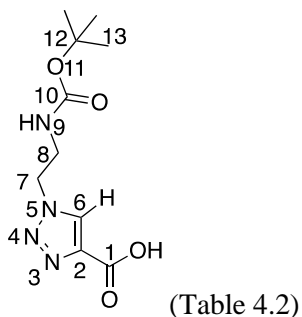
### (*R*)-3-hydroxy-2,2-dimethyl-4-oxo-4-((3-oxo-3-(propylamino)propyl)amino)butyl 1-(2-((*tert*-butoxycarbonyl)amino)ethyl)-1*H*-1,2,3-triazole-4-carboxylate, **79**



(Figure 4.5)

To a solution of 1-(2-((*tert*-butoxycarbonyl)amino)ethyl)-1*H*-1,2,3-triazole-4-carboxylic acid, **78** (0.26 g, 1.00 mmol) in DMF at 0°C, were added PyAOP (1.03 g, 1.98 mmol, 1.98 eq) and HOAt (0.14g, 1.00 mmol, 1.00 eq). To this mixture, a solution of **20** (0.26g, 0.99 mmol, 0.99 eq) and DIPEA (0.30 mL, 0.22 mmol, 0.22 eq) in DMF (8 mL) was added. The total reaction mixture was heated at 60°C whilst stirring for 15 h. The solvent was removed *in vacuo* and the crude was redissolved in EtOAc. TLC (10% MeOH in EtOAc) did not detect any new product formation. Purification by flash column chromatography (0 to 2% MeOH in EtOAc) did not give any of the product.

**1-(2-((*tert*-Butoxycarbonyl)amino)ethyl)-1*H*-1,2,3-triazole-4-carboxylic acid, **78****



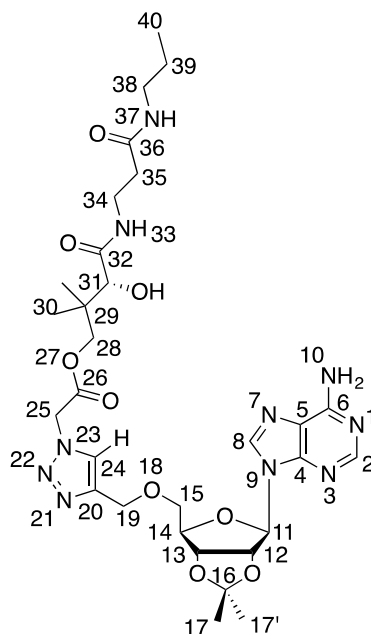
**Table 4.2**

[Reactant] M		Solvent ratio (A:B)		Catalyst loading (mol%)		Time	Yield
		<i>t</i> -BuOH	water	CuSO <sub>4</sub> ·5H <sub>2</sub> O	sodium ascorbat	stirring at r.t (h)	%
					e		
0.05	0.05	1	2	22	90	18 h	41
0.12	0.13	1	2	5.6	24	48 h	75

*Entry 1*

To a solution of propionic acid (39.5 mg, 0.55 mmol, 1.00 eq) in *t*-BuOH (1.86 mL), were added CuSO<sub>4</sub>·5H<sub>2</sub>O (3.00 mg, 0.01 mmol, 22.0 mol%) and sodium ascorbate (12.5 mg, 0.06 mmol, 90 mol%) in water (3.71 mL). The reaction mixture was stirred vigorously and *tert*-butyl (2-azidoethyl)carbamate, **77** (103 mg, 0.55 mmol) in *t*-BuOH (1.85 mL) and water (3.71 mL) were added to it. The reaction mixture was then stirred at r.t for 18 h before *t*-BuOH was removed *in vacuo*. As *t*-BuOH was removed, a white precipitate appeared. The precipitate was filtered and washed with cold water three times and dried *in vacuo* to give the product as white flakes (57 mg, 41%).

**(R)-3-hydroxy-2,2-dimethyl-4-oxo-4-((3-oxo-3-(propylamino)propyl)amino)butyl 2-(4-(((3aR,4R,6R,6aR)-6-(6-amino-9H-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-d][1,3]dioxol-4-yl)methoxy)methyl)-1H-1,2,3-triazol-1-yl)acetate, 17**



(Table 4.3)

**Table 4.3**

Entry	[Reactant] M	Solvent ratio (A:B)	Catalyst loading (mol%)	Time stirring at	Yield %		
<b>64</b>	<b>71</b>	<i>t</i> -BuOH water	CuSO <sub>4</sub> ·5H <sub>2</sub> O sodium ascorbate	r.t (h)			
1	0.1	1	1	4.8	20	48 h	X
2	0.02	1	2	6.5	23.5	48 h	X

**Entry 1**

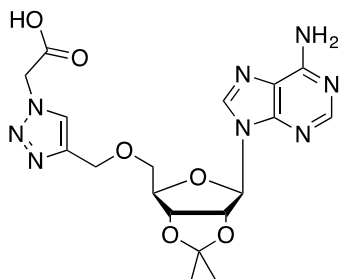
9-((3aR,4R,6R,6aR)-2,2-dimethyl-6-((prop-2-yn-1-yloxy)methyl)tetrahydrofuro[3,4-d][1,3]dioxol-4-yl)-9H-purin-6-amine, **71** (115 mg, 0.33 mmol), CuSO<sub>4</sub>·5H<sub>2</sub>O (4.00 mg, 16.0 μmol, 4.8 mol%), sodium ascorbate (13.4 mg, 68.0 μmol, 20 mol%), and (R)-3-hydroxy-2,2-dimethyl-4-oxo-4-((3-oxo-3-(propylamino)propyl)amino)butyl 2-azidoacetate, **64** (190 mg, 0.55 mmol, 1.67 eq) were dissolved in *t*-BuOH (3.1 mL) and water (3.1 mL). The reaction mixture was stirred at r.t for 48 h. LCMS did not detect any product formation.

**Entry 2**

9-((3aR,4R,6R,6aR)-2,2-dimethyl-6-((prop-2-yn-1-yloxy)methyl)tetrahydrofuro[3,4-d][1,3]dioxol-4-yl)-9H-purin-6-amine, **71** (46.2 mg, 125 μmol), CuSO<sub>4</sub>·5H<sub>2</sub>O (2.00 mg, 8.00 μmol, 6.5 mol%), sodium ascorbate (9.70 mg, 45.0 μmol, 23.5 mol%), and (R)-3-hydroxy-2,2-dimethyl-4-oxo-4-((3-oxo-3-(propylamino)propyl)amino)butyl 2-azidoacetate, **64** (48.0 mg,

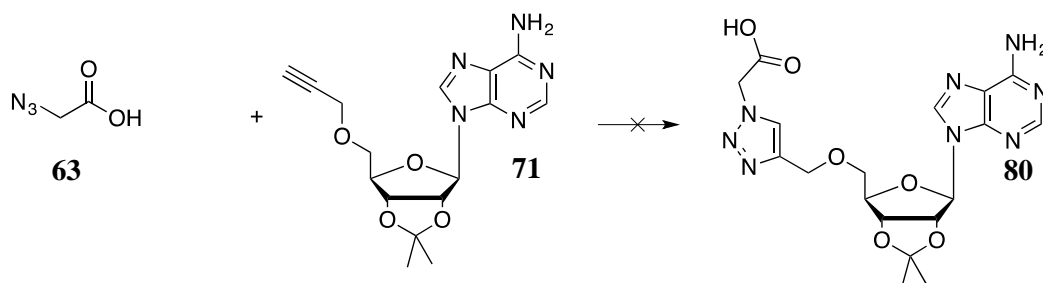
126  $\mu\text{mol}$ , 1.01 eq) were dissolved in *t*-BuOH (2.3 mL) and water (4.6 mL). The reaction mixture was stirred at r.t for 48 h. LCMS did not detect any product formation.

**2-(4-(((3*aR*,4*R*,6*R*,6*aR*)-6-(6-amino-9*H*-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)methoxy)methyl)-1*H*-1,2,3-triazol-1-yl)acetic acid, **80****



(Table 4.4)

**Table 4.4**



Entry	[Reactant] M		Solvent ratio (A:B)		Catalyst loading <sup>†</sup> (mol%)		Time (h), T (°C)	Yield %
	<b>63</b>	<b>71</b>	<i>t</i> -BuOH	water	CuSO <sub>4</sub> ·5H <sub>2</sub> O	sodium ascorbate		
1	0.06	0.05	1	1	1.8	9	48 h, r.t	X
2	0.07	0.07	1	1	0.9	9	48 h, r.t	X
3	0.06	0.05	1	1	0.9	18	48 h, r.t	X
4	0.02	0.02	1	1	1.5	7.5	1 h, microw 100°C	X

<sup>†</sup>Catalysts' stock solution in water: [CuSO<sub>4</sub>·5H<sub>2</sub>O] = [sodium ascorbate] = 0.1 M for the first 3 entries and of 0.05 M for the last entry.

*Entry 1*

9-((3*aR*,4*R*,6*R*,6*aR*)-2,2-dimethyl-6-((prop-2-yn-1-yloxy)methyl)tetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)-9*H*-purin-6-amine, **71** (51.0 mg, 0.11 mmol), along with CuSO<sub>4</sub>·5H<sub>2</sub>O (0.02 mL of 0.1 M stock solution, 2.00  $\mu\text{mol}$ , 1.8 mol%), sodium ascorbate (0.1 mL of 0.1 M stock solution, 0.01 mmol, 9 mol%), and 2-azidoacetic acid, **63** (13.0 mg, 0.12 mmol, 1.09 eq) were dissolved in *t*-BuOH (1.0 mL) and water (1.0 mL). The reaction mixture was stirred at r.t for 48 h, and monitored by LCMS over 48 h. No product was detected.

Entry 2

9-((3*aR*,4*R*,6*R*,6*aR*)-2,2-dimethyl-6-((prop-2-yn-1-yloxy)methyl)tetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)-9*H*-purin-6-amine, **71** (53.0 mg, 0.17 mmol), along with CuSO<sub>4</sub>·5H<sub>2</sub>O (0.01 mL of 0.1 M stock solution in water, 1.00 μmol, 0.9 mol%), sodium ascorbate (0.1 mL of 0.1 M stock solution in water, 0.01 mmol, 9 mol%), and 2-azidoacetic acid, **63** (15.6 mg, 0.16 mmol, 0.94 eq) were dissolved in *t*-BuOH (1.1 mL) and water (1.1 mL). The reaction mixture was stirred at r.t for 48 h, and monitored by LCMS over 48 h. No product was detected.

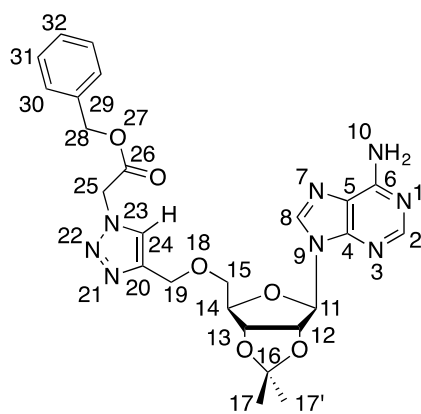
Entry 3

9-((3*aR*,4*R*,6*R*,6*aR*)-2,2-dimethyl-6-((prop-2-yn-1-yloxy)methyl)tetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)-9*H*-purin-6-amine, **71** (51.2 mg, 0.11 mmol), along with CuSO<sub>4</sub>·5H<sub>2</sub>O (0.01 mL of 0.1 M stock solution in water, 1.00 μmol, 0.9 mol%), sodium ascorbate (0.2 mL of 0.1 M stock solution in water, 0.02 mmol, 18 mol%), and 2-azidoacetic acid, **63** (14.5 mg, 0.14 mmol, 1.27 eq) were dissolved in *t*-BuOH (1.1 mL) and water (1.1 mL). The reaction mixture was stirred at r.t for 48 h, and monitored by LCMS over 48 h. No product was detected.

Entry 4

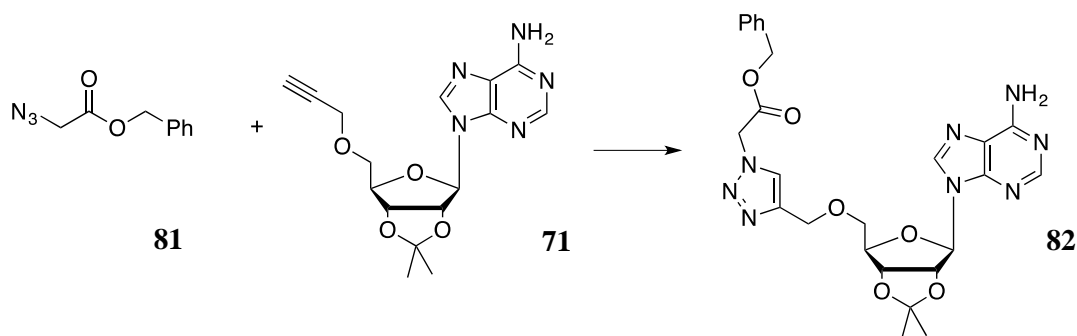
9-((3*aR*,4*R*,6*R*,6*aR*)-2,2-dimethyl-6-((prop-2-yn-1-yloxy)methyl)tetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)-9*H*-purin-6-amine, **71** (23.8 mg, 69.0 μmol), along with CuSO<sub>4</sub>·5H<sub>2</sub>O (0.02 mL of 0.05 M stock solution in water, 1.00 μmol, 1.5 mol%), sodium ascorbate (0.1 mL of 0.05 M stock solution in water, 5.00 μmol, 7.5 mol%), and 2-azidoacetic acid, **63** (7.00 mg, 69.0 μmol, 1.00 eq) were dissolved in *t*-BuOH (1.8 mL) and water (1.7 mL). The reaction mixture was stirred and heated under microwave at 100°C for 1 h, and monitored by LCMS. No product was detected. Starting materials were degraded.

**Benzyl 2-(4-(((3*aR*,4*R*,6*R*,6*aR*)-6-(6-amino-9*H*-purin-9-yl)-2,2-dimethyltetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)methoxy)methyl)-1*H*-1,2,3-triazol-1-yl)acetate, **82****



(Table 4.5 and Table 4.6)



**Table 4.5**

Entry	[Reactant] M		Solvent ratio (A:B)		Catalyst loading † (mol%)		Time (h), T (°C)	Yield %
	<b>81</b>	<b>71</b>	<i>t</i> -BuOH	water	CuSO <sub>4</sub> ·5H <sub>2</sub> O	sodium ascorbate		
1	0.02	0.02	1	1	0.75	7.5	312 h, r.t	59
2	0.02	0.02	1	1	0.75	15	20 min, 100°C microw	Trace of product in LCMS
3	0.02	0.02	1	1	0.75	15	1.3 h, 100°C microw	Trace of product in LCMS
4	0.02	0.02	1	1	0.75	15	1.3 h, 100°C, then 1 h, 150°C microw	Degrade d X
5	7 x 10 <sup>-3</sup> then 0.02	0.02	1	1	0.8	7.3	264 h, r.t ‡	25
6	0.02	0.02	1	1	0.2	1.85	192 h, r.t	56

†Catalysts' stock solution in water: [CuSO<sub>4</sub>·5H<sub>2</sub>O] = [sodium ascorbate] = 0.05 M. ‡ 24 h, r.t at 7 x 10<sup>-3</sup> M of **81** then 240 h at 0.02 M of **81**

#### Entry 1

9-((3a*R*,4*R*,6*R*,6a*R*)-2,2-dimethyl-6-((prop-2-yn-1-yloxy)methyl)tetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)-9*H*-purin-6-amine, **71** (23.8 mg, 69.0 μmol), CuSO<sub>4</sub>·5H<sub>2</sub>O (0.01 mL of 0.05 M stock solution, 0.50 μmol, 0.75 mol%), sodium ascorbate (0.1 mL of 0.05 M stock solution, 5.00 μmol, 7.5 mol%), and benzyl 2-azidoacetate, **81** (14.8 mg, 77.0 μmol, 1.10 eq) were dissolved in *t*-BuOH (1.8 mL) and water (1.8 mL). The reaction mixture was stirred at r.t for 312 h, and monitored by LCMS over the course. LCMS detected the formation of the product.

The reaction mixture was then concentrated *in vacuo* and purification by flash column chromatography (2 to 3.3% MeOH in CH<sub>2</sub>Cl<sub>2</sub>) afforded the product as a colourless oil, forming foam upon drying *in vacuo* (22 mg, 59%).

*Entry 2*

9-((3*aR*,4*R*,6*R*,6*aR*)-2,2-dimethyl-6-((prop-2-yn-1-yloxy)methyl)tetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)-9*H*-purin-6-amine, **71** (23.8 mg, 69.0 μmol), CuSO<sub>4</sub>·5H<sub>2</sub>O (0.01 mL of 0.05 M stock solution, 0.50 μmol, 0.75 mol%), sodium ascorbate (0.2 mL of 0.05 M stock solution, 10.0 μmol, 15 mol%), and benzyl 2-azidoacetate, **81** (14.8 mg, 77.0 μmol, 1.10 eq) were dissolved in *t*-BuOH (1.8 mL) and water (1.8 mL). The reaction mixture was stirred and heated under microwave at 100°C for 20 min. LCMS only detected a trace of the product.

*Entry 3*

9-((3*aR*,4*R*,6*R*,6*aR*)-2,2-dimethyl-6-((prop-2-yn-1-yloxy)methyl)tetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)-9*H*-purin-6-amine, **71** (23.8 mg, 69.0 μmol), CuSO<sub>4</sub>·5H<sub>2</sub>O (0.01 mL of 0.05 M stock solution, 0.50 μmol, 0.75 mol%), sodium ascorbate (0.2 mL of 0.05 M stock solution, 10.0 μmol, 15 mol%), and benzyl 2-azidoacetate, **81** (14.8 mg, 77.0 μmol, 1.10 eq) were dissolved in *t*-BuOH (1.8 mL) and water (1.8 mL). The reaction mixture was stirred and heated under microwave at 100°C for 80 min and monitored by LCMS in 20 min intervals. LCMS only detected a trace of the product which did not improve over the course of time.

*Entry 4*

9-((3*aR*,4*R*,6*R*,6*aR*)-2,2-dimethyl-6-((prop-2-yn-1-yloxy)methyl)tetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)-9*H*-purin-6-amine, **71** (23.8 mg, 69.0 μmol), CuSO<sub>4</sub>·5H<sub>2</sub>O (0.01 mL of 0.05 M stock solution, 0.50 μmol, 0.75 mol%), sodium ascorbate (0.2 mL of 0.05 M stock solution, 10.0 μmol, 15 mol%), and benzyl 2-azidoacetate, **81** (14.8 mg, 77.0 μmol, 1.10 eq) were dissolved in *t*-BuOH (1.8 mL) and water (1.8 mL). The reaction mixture was stirred and heated under microwave at 100°C for 80 min and at 150°C for 1 h monitored by LCMS in 20 min intervals. LCMS only detected a trace of the product when the temperature was 100°C, which did not improve over the course of time. After heating under microwave at 150°C for 1 h, starting materials were degraded.

*Entry 5*

9-((3*aR*,4*R*,6*R*,6*aR*)-2,2-dimethyl-6-((prop-2-yn-1-yloxy)methyl)tetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)-9*H*-purin-6-amine, **71** (152 mg, 0.44 mmol), CuSO<sub>4</sub>·5H<sub>2</sub>O (0.07 mL of 0.05 M stock solution, 3.50 μmol, 0.8 mol%), sodium ascorbate (0.64 mL of 0.05 M stock solution, 32.0 μmol, 7.3 mol%), and benzyl 2-azidoacetate, **81** (30.0 mg, 0.16 μmol, 0.36 eq) were dissolved in *t*-BuOH (11 mL) and water (11 mL). The reaction mixture was stirred at r.t for 24

h. LCMS detected the formation of the product. Benzyl 2-azidoacetate, **81** (54.0 mg, 0.28  $\mu\text{mol}$ , 0.64 eq) was added further to the stirring reaction mixture. The reaction mixture was then stirred at r.t for a further 240 h. LCMS indicated approximately 45% completion of the reaction. After that, it was concentrated *in vacuo* and purification by flash column chromatography (2 to 4% MeOH in  $\text{CH}_2\text{Cl}_2$ ) afforded the product as a colourless oil, forming foam upon drying *in vacuo* (60 mg, 25%).

*Entry 6*

9-((3*aR*,4*R*,6*R*,6*aR*)-2,2-dimethyl-6-((prop-2-yn-1-yloxy)methyl)tetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)-9*H*-purin-6-amine, **71** (416 mg, 1.20 mmol),  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$  (0.19 mL of 0.05 M stock solution, 9.50  $\mu\text{mol}$ , 0.2 mol%), sodium ascorbate (1.75 mL of 0.05 M stock solution, 87.5  $\mu\text{mol}$ , 1.85 mol%), and benzyl 2-azidoacetate, **81** (229 mg, 1.20 mmol, 1.00 eq) were dissolved in *t*-BuOH (30 mL) and water (30 mL). The reaction mixture was stirred at r.t for 192 h. LCMS indicated approximately 61% completion of the reaction. After that, it was concentrated *in vacuo* and purification by flash column chromatography (2 to 4% MeOH in  $\text{CH}_2\text{Cl}_2$ ) afforded the product as a colourless oil, forming foam upon drying *in vacuo* (220 mg, 56%).

**Table 4.6**

Entry	[Reactant] M		Solvent ratio (A:B)		Catalyst loading <sup>†</sup> (mol%)		Time (h), r.t	Completion %
	<b>81</b>	<b>71</b>	<i>t</i> -BuOH	water	$\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$	sodium ascorbate		
1	0.06	0.06	1	1	0.8	7.5	26 h	90
2	0.15	0.15	1	1	0.8	7.5	26 h	80
3	0.06	0.06	1	1	0.8	7.5	24 h	96

<sup>†</sup>Catalysts' stock solution:  $[\text{CuSO}_4 \cdot 5\text{H}_2\text{O}] = [\text{sodium ascorbate}] = 0.05 \text{ M}$ .

*Entry 1*

9-((3*aR*,4*R*,6*R*,6*aR*)-2,2-dimethyl-6-((prop-2-yn-1-yloxy)methyl)tetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)-9*H*-purin-6-amine, **71** (100 mg, 0.29 mmol),  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$  (45  $\mu\text{L}$  of 0.05 M stock solution, 2.25  $\mu\text{mol}$ , 0.8 mol%), sodium ascorbate (0.42 mL of 0.05 M stock solution, 21.0  $\mu\text{mol}$ , 7.5 mol%), and benzyl 2-azidoacetate, **81** (55.9 mg, 0.29 mmol, 1.00 eq) were dissolved in *t*-BuOH (2.4 mL) and water (2.4 mL). After stirring at r.t for 26 h, LCMS indicated 90% completion of the reaction.

*Entry 2*

9-((3*aR*,4*R*,6*R*,6*aR*)-2,2-dimethyl-6-((prop-2-yn-1-yloxy)methyl)tetrahydrofuro[3,4-*d*][1,3]dioxol-4-yl)-9*H*-purin-6-amine, **71** (101 mg, 0.29 mmol), CuSO<sub>4</sub>·5H<sub>2</sub>O (45 μL of 0.05 M stock solution, 2.25 μmol, 0.8 mol%), sodium ascorbate (0.42 mL of 0.05 M stock solution, 21.0 μmol, 7.5 mol%), and benzyl 2-azidoacetate, **81** (57.0 mg, 0.30 mmol, 1.03 eq) were dissolved in *t*-BuOH (1.0 mL) and water (1.0 mL). After stirring at r.t for 26 h, LCMS indicated 80% completion of the reaction.

## 7. Appendix

### *Appendix 2.1. PDBs used in structural alignment of protein kinases*

PDB (AURKA)				PDB (AURKB)			PDB (PKA)				
1MQ4	1OL5	1OL7	2BMC	2BFX	2BFY	4AF3	1APM	1ATP	1BKX	1BX6	1CDK
2C6E	2DWB	2J4Z	2NP8	4B8L	4B8M	4C2V	1CME	1CTP	1FMO	1J3H	1JBP
2WTV	2WTW	2X6D	2X6E		4C2W		1JLU	1L3R	1Q24	1Q61	1Q62
2X81	2XRU	3E5A	3EFW				1Q8T	1Q8U	1Q8W	1RDQ	1RE8
3H0Y	3H0Z	3H0Z	3H10				1REJ	1REK	1SMH	1STC	1SVE
3HA6	3K5U	3LAU	3M11				1SVG	1SVH	1SZM	1XH4	1XH5
3MYG	3NRM	3O5O	3O51				1XH6	1XH7	1XH8	1XH9	1XHA
3R21	3UO4	3UOJ	3UOL				1YDR	1YDS	1YDT	2C1A	2C1B
3UP2	3UP7	3VAP	3W18				2ERZ	2F7E	2F7X	2GFC	2GNF
3W2C	4B0G	4BN1	4BYJ				2GNG	2GNH	2GNI	2GNJ	2GNL
4DEE	4DHF	4J8M	4JBO				2GU8	2JDS	2JDT	2JDV	2OH0
4JBP	4JBQ	4PRJ	4UYN				2OJF	2QCS	2QUR	2QVS	2UVX
4UZD	4UZH						2UVY	2UVZ	2UW0	2UW3	2UW4
							2UW5	2UW6	2UW7	2UW8	2UZT
							2UZU	2UZV	2UZW	2VNW	2VNY
							2VO0	2VO3	2VO6	2VO7	3AG9
							3AGL	3AGM	3AMA	3AMB	3BWJ
							3DND	3DNC	3E8C	3E8E	3FHI
							3FJQ	3IDB	3KKV	3L9L	3L9M
							3L9N	3NNJ	3NX8	3OOG	3OVV
							3OW3	3OWP	3OXT	3P0M	3POD
							3PVB	3QAL	3QAM	3TNP	3VQH
							3ZO1	3ZO2	3ZO3	3ZO4	4AE6
							4AE9	4AXA	4C33	4C34	4C35
							4C36	4C37	4C38	4DFX	4DFZ
							4DG0	4DG2	4DG3	4DH1	4DH3
							4DH5	4DH7	4DH8	4HPT	4HPU
							4IAC	4IAD	4IAF	4IAI	4IAK
							4IAY	4IAZ	4IB0	4IB1	4IB3
							4IE9	4IJ9	4O21	4O22	4WIH

*Appendix 2.2. Comparison of the B-factors of the hinge region and activation loop of 1OL7 and 4BNI*

**1OL7**

<b>ATOM</b>	<b>NUMBER</b>	<b>RESIDUE</b>	<b>B-FACTOR</b>	<b>ELEMENT</b>
ATOM	1217	N TRP A 277	40.57	N
ATOM	1218	CA TRP A 277	36.62	C
ATOM	1219	C TRP A 277	35.74	C
ATOM	1220	O TRP A 277	35.61	O
ATOM	1221	CB TRP A 277	42.09	C
ATOM	1222	CG TRP A 277	50.97	C
ATOM	1223	CD1 TRP A 277	54.30	C
ATOM	1224	CD2 TRP A 277	58.62	C
ATOM	1225	NE1 TRP A 277	59.86	N
ATOM	1226	CE2 TRP A 277	58.91	C
ATOM	1227	CE3 TRP A 277	59.93	C
ATOM	1228	CZ2 TRP A 277	61.19	C
ATOM	1229	CZ3 TRP A 277	61.29	C
ATOM	1230	CH2 TRP A 277	63.73	C
ATOM	1231	N SER A 278	44.94	N
ATOM	1232	CA SER A 278	46.77	C
ATOM	1233	C SER A 278	50.47	C
ATOM	1234	O SER A 278	54.17	O
ATOM	1235	CB SER A 278	42.80	C
ATOM	1236	OG SER A 278	54.88	O
ATOM	1237	N VAL A 279	51.50	N
ATOM	1238	CA VAL A 279	57.94	C
ATOM	1239	C VAL A 279	59.15	C
ATOM	1240	O VAL A 279	67.05	O
ATOM	1241	CB VAL A 279	44.98	C
ATOM	1242	CG1 VAL A 279	43.31	C
ATOM	1243	CG2 VAL A 279	57.04	C
ATOM	1244	N HIS A 280	63.81	N
ATOM	1245	CA HIS A 280	67.43	C
ATOM	1246	C HIS A 280	70.50	C
ATOM	1247	O HIS A 280	82.51	O
ATOM	1248	CB HIS A 280	66.79	C
ATOM	1249	CG HIS A 280	72.05	C
ATOM	1250	ND1 HIS A 280	73.34	N
ATOM	1251	CD2 HIS A 280	75.11	C
ATOM	1252	CE1 HIS A 280	80.03	C
ATOM	1253	NE2 HIS A 280	83.14	N
ATOM	1254	N ALA A 281	68.02	N
ATOM	1255	CA ALA A 281	68.33	C
ATOM	1256	C ALA A 281	68.64	C
ATOM	1257	O ALA A 281	67.99	O
ATOM	1258	CB ALA A 281	60.98	C
ATOM	1259	N PRO A 282	70.43	N
ATOM	1260	CA PRO A 282	67.06	C
ATOM	1261	C PRO A 282	69.62	C
ATOM	1262	O PRO A 282	80.00	O
ATOM	1263	CB PRO A 282	67.65	C
ATOM	1264	CG PRO A 282	63.34	C
ATOM	1265	CD PRO A 282	64.78	C
ATOM	1266	N SER A 283	65.02	N

ATOM	1267	CA	SER	A	283	59.53	C
ATOM	1268	C	SER	A	283	55.99	C
ATOM	1269	O	SER	A	283	69.93	O
ATOM	1270	CB	SER	A	283	55.46	C
ATOM	1271	OG	SER	A	283	55.84	O
ATOM	1272	N	SER	A	284	60.56	N
ATOM	1273	CA	SER	A	284	63.49	C
ATOM	1274	C	SER	A	284	63.54	C
ATOM	1275	O	SER	A	284	57.98	O
ATOM	1276	CB	SER	A	284	63.76	C
ATOM	1277	OG	SER	A	284	77.24	O
ATOM	1278	N	ARG	A	285	64.16	N
ATOM	1279	CA	ARG	A	285	67.66	C
ATOM	1280	C	ARG	A	285	68.61	C
ATOM	1281	O	ARG	A	285	72.84	O
ATOM	1282	CB	ARG	A	285	62.91	C
ATOM	1283	N	ARG	A	286	71.24	N
ATOM	1284	CA	ARG	A	286	75.06	C
ATOM	1285	C	ARG	A	286	79.18	C
ATOM	1286	O	ARG	A	286	80.60	O
ATOM	1287	CB	ARG	A	286	73.44	C
ATOM	1288	CG	ARG	A	286	71.75	C
ATOM	1289	CD	ARG	A	286	72.37	C
ATOM	1290	NE	ARG	A	286	72.92	N
ATOM	1291	CZ	ARG	A	286	77.01	C
ATOM	1292	NH1	ARG	A	286	80.80	N
ATOM	1293	NH2	ARG	A	286	73.00	N
HETATM	1294	N	TPO	A	287	82.59	N
HETATM	1295	CA	TPO	A	287	81.36	C
HETATM	1296	CB	TPO	A	287	85.77	C
HETATM	1297	CG2	TPO	A	287	78.62	C
HETATM	1298	OG1	TPO	A	287	96.25	O
HETATM	1299	P	TPO	A	287	94.10	P
HETATM	1300	O1P	TPO	A	287	105.80	O
HETATM	1301	O2P	TPO	A	287	103.14	O
HETATM	1302	O3P	TPO	A	287	103.19	O
HETATM	1303	C	TPO	A	287	84.40	C
HETATM	1304	O	TPO	A	287	82.15	O
HETATM	1305	N	TPO	A	288	86.52	N
HETATM	1306	CA	TPO	A	288	85.55	C
HETATM	1307	CB	TPO	A	288	87.59	C
HETATM	1308	CG2	TPO	A	288	82.26	C
HETATM	1309	OG1	TPO	A	288	93.55	O
HETATM	1310	P	TPO	A	288	87.77	P
HETATM	1311	O1P	TPO	A	288	91.79	O
HETATM	1312	O2P	TPO	A	288	93.35	O
HETATM	1313	O3P	TPO	A	288	95.16	O
HETATM	1314	C	TPO	A	288	89.28	C
HETATM	1315	O	TPO	A	288	86.14	O
ATOM	1316	N	LEU	A	289	87.01	N
ATOM	1317	CA	LEU	A	289	88.25	C
ATOM	1318	C	LEU	A	289	87.34	C
ATOM	1319	O	LEU	A	289	90.27	O
ATOM	1320	CB	LEU	A	289	78.91	C
ATOM	1321	N	CYS	A	290	89.51	N
ATOM	1322	CA	CYS	A	290	89.73	C
ATOM	1323	C	CYS	A	290	90.22	C
ATOM	1324	O	CYS	A	290	91.11	O

ATOM	1325	CB	CYS	A	290	80.05	C
ATOM	1326	N	GLY	A	291	90.88	N
ATOM	1327	CA	GLY	A	291	84.36	C
ATOM	1328	C	GLY	A	291	80.67	C
ATOM	1329	O	GLY	A	291	72.19	O
ATOM	1330	N	THR	A	292	72.87	N
ATOM	1331	CA	THR	A	292	68.66	C
ATOM	1332	C	THR	A	292	69.67	C
ATOM	1333	O	THR	A	292	81.32	O
ATOM	1334	CB	THR	A	292	64.26	C
ATOM	1335	OG1	THR	A	292	62.83	O
ATOM	1336	CG2	THR	A	292	66.76	C
ATOM	1337	N	LEU	A	293	65.76	N
ATOM	1338	CA	LEU	A	293	57.85	C
ATOM	1339	C	LEU	A	293	54.37	C
ATOM	1340	O	LEU	A	293	59.11	O
ATOM	1341	CB	LEU	A	293	58.26	C
ATOM	1342	CG	LEU	A	293	59.75	C
ATOM	1343	CD1	LEU	A	293	56.85	C
ATOM	1344	CD2	LEU	A	293	55.33	C
ATOM	1345	N	ASP	A	294	45.72	N
ATOM	1346	CA	ASP	A	294	46.04	C
ATOM	1347	C	ASP	A	294	42.63	C
ATOM	1348	O	ASP	A	294	40.46	O
ATOM	1349	CB	ASP	A	294	47.18	C
ATOM	1350	CG	ASP	A	294	53.87	C
ATOM	1351	OD1	ASP	A	294	53.07	O
ATOM	1352	OD2	ASP	A	294	44.18	O

#### 4BN1

ATOM	NUMBER	RESIDUE	B-FACTOR	ELEMENT
ATOM	1222	N TRP A 277	50.36	N
ATOM	1223	CA TRP A 277	55.42	C
ATOM	1224	C TRP A 277	51.58	C
ATOM	1225	O TRP A 277	51.76	O
ATOM	1226	CB TRP A 277	50.87	C
ATOM	1227	CG TRP A 277	58.49	C
ATOM	1228	CD1 TRP A 277	63.00	C
ATOM	1229	CD2 TRP A 277	67.07	C
ATOM	1230	NE1 TRP A 277	63.47	N
ATOM	1231	CE2 TRP A 277	73.15	C
ATOM	1232	CE3 TRP A 277	65.12	C
ATOM	1233	CZ2 TRP A 277	77.08	C
ATOM	1234	CZ3 TRP A 277	70.17	C
ATOM	1235	CH2 TRP A 277	70.70	C
ATOM	1236	N SER A 278	49.59	N
ATOM	1237	CA SER A 278	55.05	C
ATOM	1238	C SER A 278	56.32	C
ATOM	1239	O SER A 278	54.31	O
ATOM	1240	CB SER A 278	47.01	C
ATOM	1241	OG SER A 278	67.87	O
ATOM	1242	N VAL A 279	50.55	N
ATOM	1243	CA VAL A 279	56.77	C
ATOM	1244	C VAL A 279	62.31	C
ATOM	1245	O VAL A 279	54.29	O

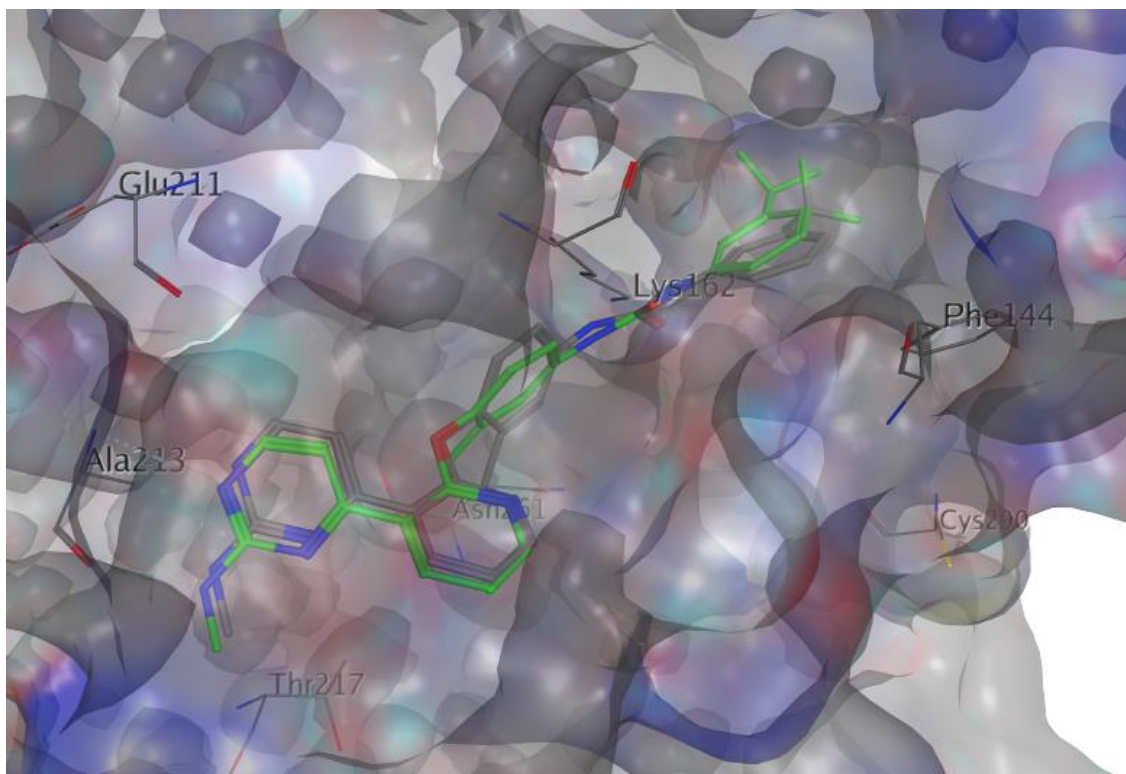


ATOM	1246	CB	VAL	A	279	56.82	C
ATOM	1247	CG1	VAL	A	279	50.42	C
ATOM	1248	CG2	VAL	A	279	52.79	C
ATOM	1249	N	HIS	A	280	65.48	N
ATOM	1250	CA	HIS	A	280	67.21	C
ATOM	1251	C	HIS	A	280	59.30	C
ATOM	1252	O	HIS	A	280	67.63	O
ATOM	1253	CB	HIS	A	280	69.35	C
ATOM	1254	CG	HIS	A	280	74.61	C
ATOM	1255	ND1	HIS	A	280	68.93	N
ATOM	1256	CD2	HIS	A	280	74.09	C
ATOM	1257	CE1	HIS	A	280	71.38	C
ATOM	1258	NE2	HIS	A	280	74.23	N
ATOM	1259	N	ALA	A	281	57.47	N
ATOM	1260	CA	ALA	A	281	66.72	C
ATOM	1261	C	ALA	A	281	73.42	C
ATOM	1262	O	ALA	A	281	71.50	O
ATOM	1263	CB	ALA	A	281	64.66	C
ATOM	1264	N	PRO	A	282	71.56	N
ATOM	1265	CA	PRO	A	282	65.81	C
ATOM	1266	C	PRO	A	282	68.41	C
ATOM	1267	O	PRO	A	282	60.30	O
ATOM	1268	CB	PRO	A	282	69.11	C
ATOM	1269	CG	PRO	A	282	74.40	C
ATOM	1270	CD	PRO	A	282	62.45	C
ATOM	1271	N	SER	A	283	71.62	N
ATOM	1272	CA	SER	A	283	62.93	C
ATOM	1273	C	SER	A	283	59.42	C
ATOM	1274	O	SER	A	283	59.19	O
ATOM	1275	CB	SER	A	283	57.85	C
ATOM	1276	OG	SER	A	283	70.25	O
ATOM	1277	N	SER	A	284	69.23	N
ATOM	1278	CA	SER	A	284	75.78	C
ATOM	1279	C	SER	A	284	85.22	C
ATOM	1280	O	SER	A	284	90.86	O
ATOM	1281	CB	SER	A	284	77.25	C
ATOM	1282	OG	SER	A	284	82.32	O
ATOM	1283	N	ARG	A	285	94.04	N
ATOM	1284	CA	ARG	A	285	93.40	C
ATOM	1285	C	ARG	A	285	100.10	C
ATOM	1286	O	ARG	A	285	114.72	O
ATOM	1287	CB	ARG	A	285	89.67	C
ATOM	1288	N	ARG	A	286	93.25	N
ATOM	1289	CA	ARG	A	286	93.98	C
ATOM	1290	C	ARG	A	286	102.38	C
ATOM	1291	O	ARG	A	286	102.50	O
ATOM	1292	CB	ARG	A	286	77.66	C
HETATM	1293	N	TPO	A	287	106.03	N
HETATM	1294	CA	TPO	A	287	94.54	C
HETATM	1295	CB	TPO	A	287	92.06	C
HETATM	1296	CG2	TPO	A	287	95.97	C
HETATM	1297	OG1	TPO	A	287	83.93	O
HETATM	1298	P	TPO	A	287	77.52	P
HETATM	1299	O1P	TPO	A	287	71.15	O
HETATM	1300	O2P	TPO	A	287	73.40	O
HETATM	1301	O3P	TPO	A	287	98.95	O
HETATM	1302	C	TPO	A	287	101.03	C
HETATM	1303	O	TPO	A	287	112.16	O

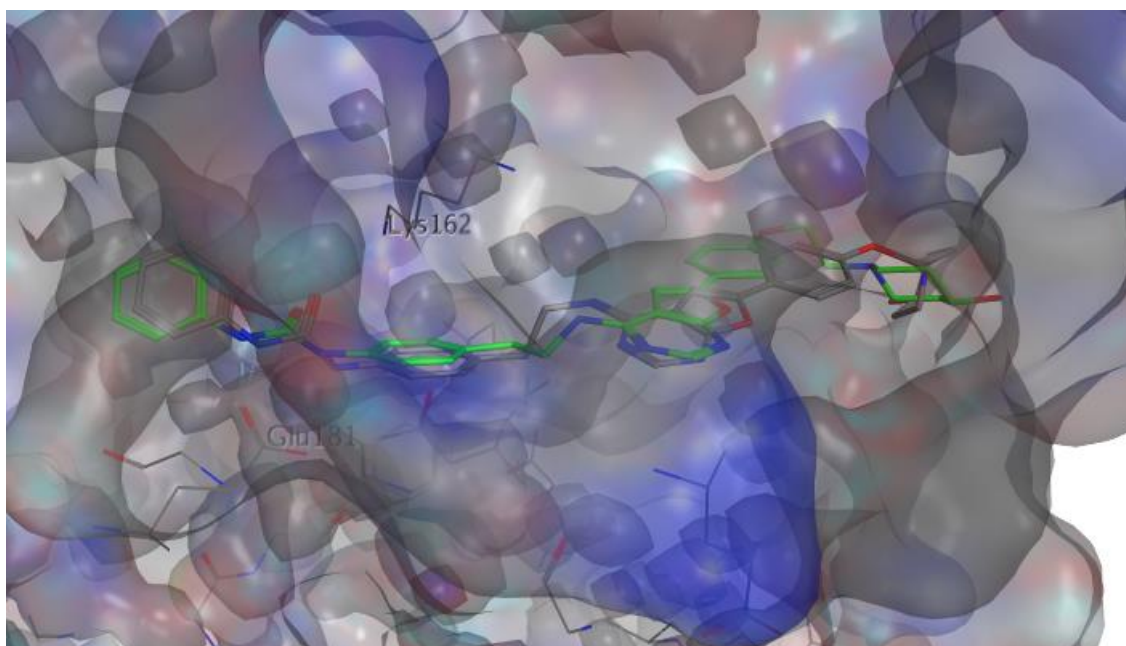
HETATM	1304	N	TPO	A	288	100.37	N
HETATM	1305	CA	TPO	A	288	109.47	C
HETATM	1306	CB	TPO	A	288	120.35	C
HETATM	1307	CG2	TPO	A	288	121.25	C
HETATM	1308	OG1	TPO	A	288	133.23	O
HETATM	1309	P	TPO	A	288	153.57	P
HETATM	1310	O1P	TPO	A	288	117.84	O
HETATM	1311	O2P	TPO	A	288	133.44	O
HETATM	1312	O3P	TPO	A	288	133.59	O
HETATM	1313	C	TPO	A	288	110.43	C
HETATM	1314	O	TPO	A	288	112.01	O
ATOM	1315	N	LEU	A	289	105.94	N
ATOM	1316	CA	LEU	A	289	106.99	C
ATOM	1317	C	LEU	A	289	109.44	C
ATOM	1318	O	LEU	A	289	103.15	O
ATOM	1319	CB	LEU	A	289	109.87	C
ATOM	1320	CG	LEU	A	289	108.24	C
ATOM	1321	CD1	LEU	A	289	92.63	C
ATOM	1322	CD2	LEU	A	289	90.67	C
ATOM	1323	N	CYS	A	290	108.82	N
ATOM	1324	CA	CYS	A	290	108.73	C
ATOM	1325	C	CYS	A	290	104.71	C
ATOM	1326	O	CYS	A	290	102.72	O
ATOM	1327	CB	CYS	A	290	92.96	C
ATOM	1328	SG	CYS	A	290	135.29	S
ATOM	1329	N	GLY	A	291	89.23	N
ATOM	1330	CA	GLY	A	291	87.52	C
ATOM	1331	C	GLY	A	291	75.64	C
ATOM	1332	O	GLY	A	291	51.20	O
ATOM	1333	N	THR	A	292	72.14	N
ATOM	1334	CA	THR	A	292	61.07	C
ATOM	1335	C	THR	A	292	52.72	C
ATOM	1336	O	THR	A	292	60.68	O
ATOM	1337	CB	THR	A	292	70.16	C
ATOM	1338	OG1	THR	A	292	61.93	O
ATOM	1339	CG2	THR	A	292	54.62	C
ATOM	1340	N	LEU	A	293	52.64	N
ATOM	1341	CA	LEU	A	293	42.61	C
ATOM	1342	C	LEU	A	293	44.20	C
ATOM	1343	O	LEU	A	293	42.90	O
ATOM	1344	CB	LEU	A	293	41.82	C
ATOM	1345	CG	LEU	A	293	54.18	C
ATOM	1346	CD1	LEU	A	293	48.52	C
ATOM	1347	CD2	LEU	A	293	46.12	C
ATOM	1348	N	ASP	A	294	43.69	N
ATOM	1349	CA	ASP	A	294	35.96	C
ATOM	1350	C	ASP	A	294	43.43	C
ATOM	1351	O	ASP	A	294	44.92	O
ATOM	1352	CB	ASP	A	294	40.14	C
ATOM	1353	CG	ASP	A	294	49.06	C
ATOM	1354	OD1	ASP	A	294	47.38	O
ATOM	1355	OD2	ASP	A	294	55.69	O

### Appendix 2.3. Validation of CHEMPLP as the docking function

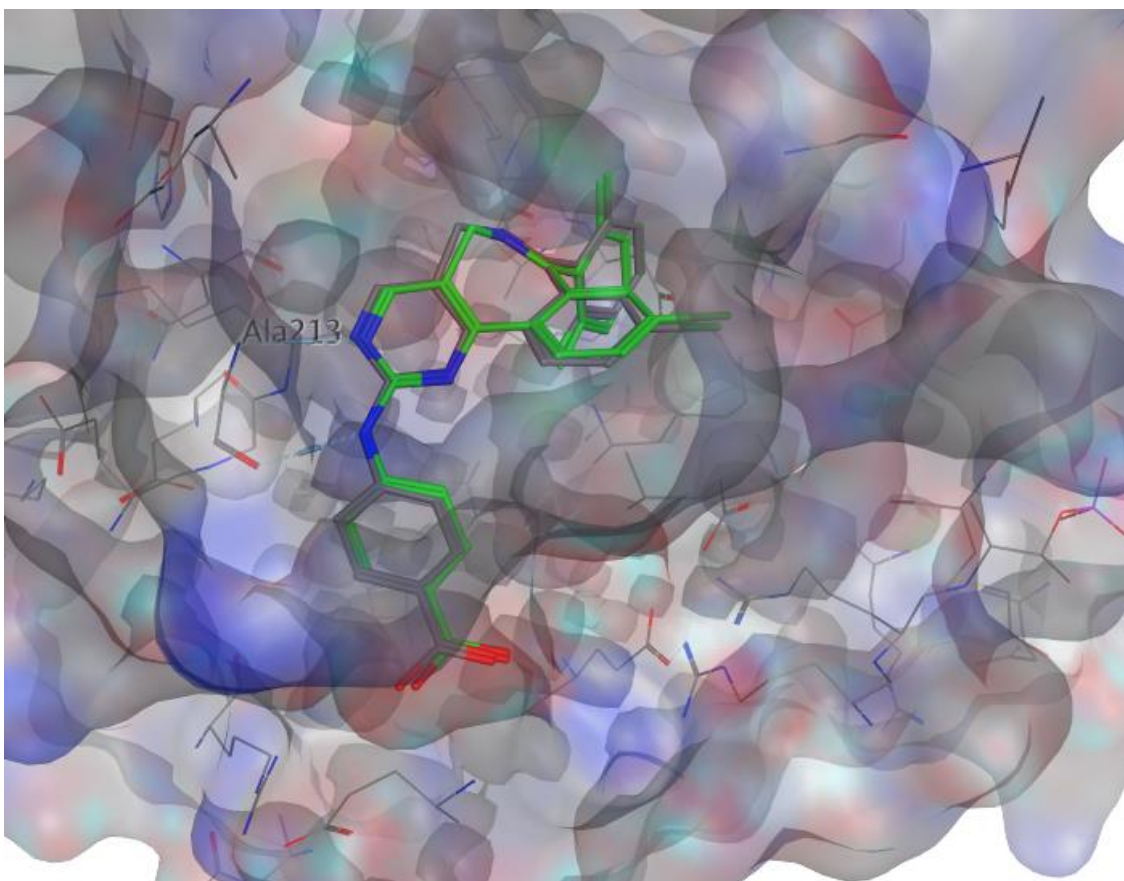
CHEMPLP was able to reproduce the conformations of the ligands AK8, YPH and ZZL as in the original crystal structures: 3EFW, 4JBP and 2WTV respectively.



**Appendix 2.3. a)** Re-evaluation of 3EFW. Ligand: AK8 re-docking result. AK8 crystallised structure (C atoms in grey), AK8 solution generated by CHEMPLP (C atoms in green). H-bonds (dashed lines).

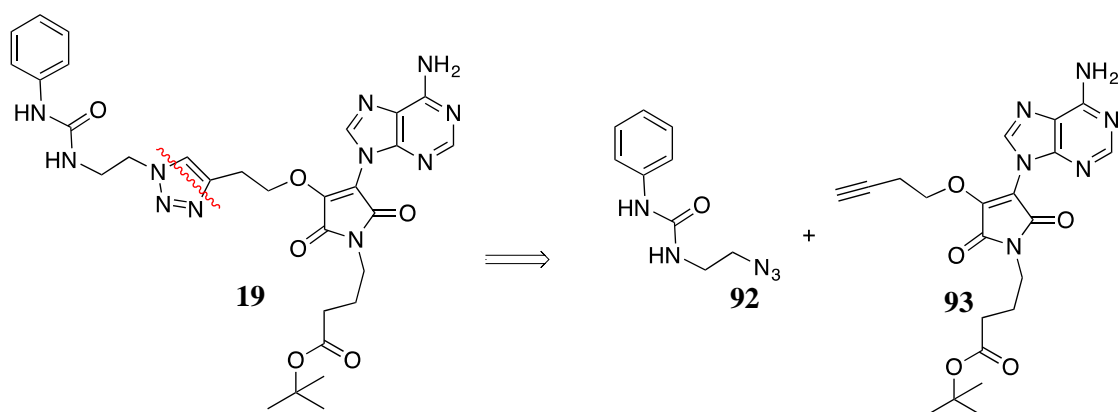


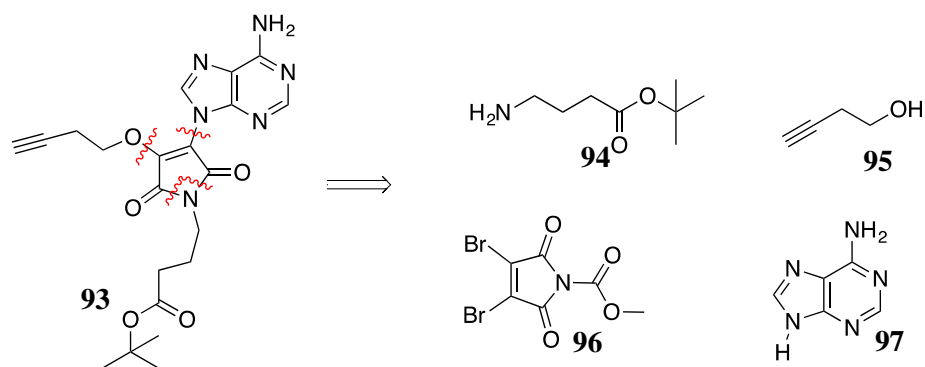
**Appendix 2.3. b)** Re-evaluation of 4JBP. Ligand: YPH re-docking result. YPH crystallised structure (C atoms in grey), YPH solution generated by CHEMPLP (C atoms in green). H-bonds (dashed lines).



**Appendix 2.3.** c) Re-evaluation of 2WTV. Ligand: ZZZL re-docking result. ZZZL crystallised structure (C atoms in grey), ZZZL solution generated by CHEMPLP (C atoms in green). H-bonds (dashed lines).

Appendix 4.1. Retrosynthesis of compound **19**

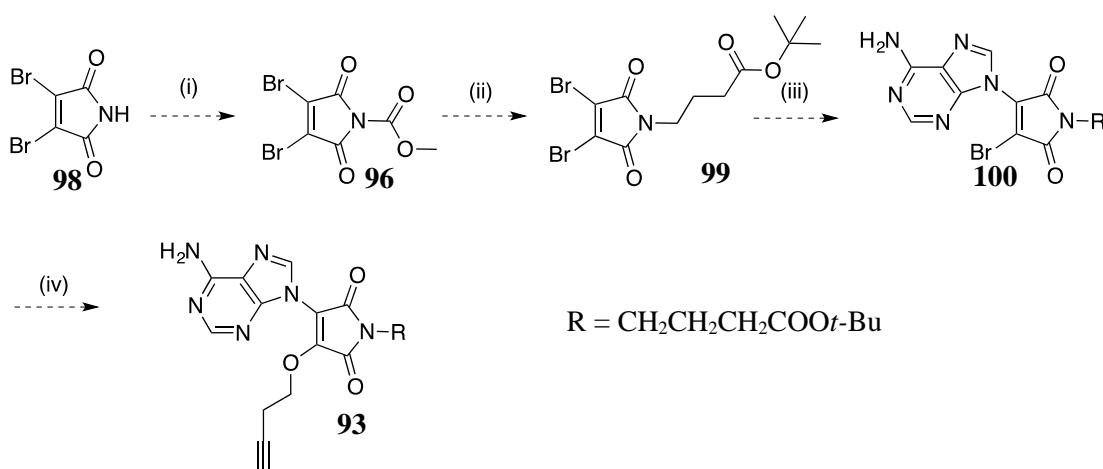




**Figure 7.1.** Retrosynthesis of **19**

Compound **19** may be synthesised from azide **92** and alkyne **93**. The synthesis of azide **92** is similar to that of 2-azido-*N*-(2-(3-phenylureido)ethyl)acetamide, **86**. It was suggested that synthesis of alkyne **93** might be completed from *N*-functionalised dibromomaleimide.<sup>177</sup> Since the *N*-functionalised dithiophenolmaleimides proved more challenging and required subtle tuning of the reactions, the dibromomaleimide variants were chosen.<sup>177</sup>

As -OR nucleophile may add twice to the same place or to an already substituted end, reaction that involves but-3-yn-1-ol as the nucleophile can be performed carefully in mild conditions (mild base) and after the reaction involving adenine as the nucleophile.<sup>178</sup> Adenine may be replaced by other amine moieties present in kinase inhibitors.



**Figure 7.2.** Proposed synthesis of **93**. (i)  $\text{ClCO}_2\text{Me}$ , NMM; (ii) *tert*-butyl 4-aminobutanoate; (iii) (protected-NH<sub>2</sub>)-adenine, base; (iv) But-3-yn-1-ol, base.

## 8. References

- <sup>1</sup> a) *Cancer: basic science and clinical aspects*, Almeida, C. A. and Barry, S. A., **2010**, John Wiley & Sons, UK; b) *Cancer Biology 3<sup>rd</sup> Edition*, King, R. J. B. and Robins, M. W., **2006**, Pearson Education Limited, Edinburgh Gate, Harlow, Essex CM20 2JE, England; c) *The biology of cancer 2<sup>nd</sup> Edition* (ed. Gabriel, J.), **2007**, John Wiley & Sons, The Atrium Southern Gate, Chichester, West Sussex, PO19 8SQ, England.
- <sup>2</sup> Cancer Research UK, 2014: <http://www.cancerresearchuk.org/health-professional/cancer-statistics>.
- <sup>3</sup> Perez de Castro, I., de Carcer, G., Malumbres, M., *Carcinogenesis* **2007**, *28*, 899-912.
- <sup>4</sup> a) Nigg, E. A. and Stearns, T., *Nat. Cell Biol.* **2011**, *13*, 1154-1160; b) Hardy, P. A. and Zacharias, H., *Cell Biol. Intl.* **2005**, *29*, 983-992.
- <sup>5</sup> Hanahan, D. and Weinberg, R. A., *Cell* **2011**, *144*, 646-674.
- <sup>6</sup> Karnoub, A. E. and Weinberg, R. A., *Breast Dis.* **2006-2007**, *26*, 75-85.
- <sup>7</sup> Hanahan, D., Weinberg, R. A., *Cell* **2000**, *100*, 57-70.
- <sup>8</sup> Burkhardt, D. L. and Sage, J., *Nat. Rev. Cancer* **2008**, *8*, 671-682.
- <sup>9</sup> Fabbro, D., Cowan-Jacob, S. W., and Moebitz, H., *Br. J. Pharm.* **2015**, *172*, 2675-2700.
- <sup>10</sup> Fleuren, E. D. G., Zhang, L., Wu, J., Daly, R. J., *Nat. Rev. Cancer* **2016**, *16*, 83-98.
- <sup>11</sup> Ubersax, J. A. and Ferrell, J. E. Jr, *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 530-541.
- <sup>12</sup> Manning, G., Whyte, D. B., Martinez, R., Hunter, T., Sudarsanam, S., *Science* **2002**, *298*, 1912-1934.
- <sup>13</sup> a) Kannan, N. and Taylor, S. S., *Cell* **2008**, *133*, 204-205; b) Rajakulendran. T. and Sicheri, F., *Sci. Signal.* **2010**, *3*, 8.
- <sup>14</sup> Chen, M., Cheng, A., Candotti, F., Zhou, Y. J., Hymel, A., Fasth, A., Notarangelo, L. D., O'Shea, J. J., *Mol. Cell. Biol.* **2000**, *20*, 947-956.
- <sup>15</sup> Futreal, P. A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., Stratton, M. R., *Nat. Rev. Cancer* **2004**, *4*, 177-183.
- <sup>16</sup> Kan, Z., Jaiswal, B. S., Stinson, J., Janakiraman, V., Bhatt, D., Stern, H. M., Yue, P., Haverty, P. M., Bourgon, R., Zheng, J., Moorhead, M., Chaudhuri, S., Tomsho, L. P., Peters, B. A., Pujara, K., Cordes, S., Davis, D. P., Carlton, V. E., Yuan, W., Li, L., Wang, W., Eigenbrot, C., Kaminker, J. S., Eberhard, D. A., Waring, P., Schuster, S. C., Modrusan, Z., Zhang, Z., Stokoe, D., de Sauvage, F. J., Faham, M., Seshagiri, S., *Nature* **2010**, *466*, 869-873.
- <sup>17</sup> a) Dees, N. D., Zhang, Q., Kandoth, C., Wendl, M. C., Schierding, W., Koboldt, D. C., Mooney, T. B., Callaway, M. B., Dooling, D., Mardis, E. R., Wilson, R. K., Ding, L., *Genome Res.* **2012**, *22*, 1589-1598; b) Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., Carter, S. L., Stewart, C., Mermel, C. H., Roberts, S. A., Kiezun, A., Hammerman, P. S., McKenna, A., Drier, Y., Zou, L., Ramos, A. H., Pugh, T. J., Stransky, N., Helman, E., Kim, J., Sougnez, C., Ambrogio, L., Nickerson, E., Shefler, E., Cortés, M. L., Auclair, D., Saksena, G., Voet, D., Noble, M., DiCara, D., Lin, P., Lichtenstein, L., Heiman, D.

I., Fennell, T., Imielinski, M., Hernandez, B., Hodis, E., Baca, S., Dulak, A. M., Lohr, J., Landau, D. A., Wu, C. J., Melendez-Zajgla, J., Hidalgo-Miranda, A., Koren, A., McCarroll, S. A., Mora, J., Lee, R. S., Crompton, B., Onofrio, R., Parkin, M., Winckler, W., Ardlie, K., Gabriel, S. B., Roberts, C. W., Biegel, J. A., Stegmaier, K., Bass, A. J., Garraway, L. A., Meyerson, M., Golub, T. R., Gordenin, D. A., Sunyaev, S., Lander, E. S., Getz, G., *Nature* **2013**, *499*, 214-218; c) Lawrence, M. S., Stojanov, P., Mermel, C. H., Robinson, J. T., Garraway, L. A., Golub, T. R., Meyerson, M., Gabriel, S. B., Lander, E. S., Getz, G., *Nature* **2014**, *505*, 495-501.

<sup>18</sup> Zhang, J., Yang, P. L., Gray, N. S., *Nat. Rev.* **2009**, *9*, 28-39.

<sup>19</sup> PDB: 1OL7 and 1OL5. Bayliss, R., Sardon, T., Vernos, I., Conti, E., *Mol. Cell.* **2003**, *12*, 851-862.

<sup>20</sup> a) Bhattacharyya, R. P., Remenyi, A., Yeh, B. J. & Lim, W. A. *Annu. Rev. Biochem.* **2006**, *75*, 655-680; b) Remenyi, A., Good, M. C. and Lim, W. A. *Curr. Opin. Struct. Biol.* **2006**, *16*, 676-685.

<sup>21</sup> a) Biondi, R. M., Cheung, P. C. F., Casamayor, A., Deak, M., Currie, R. A., Alessi, D. R., *EMBO J.* **2002**, *19*, 979-988; b) Frödin, M., L. Antal, T. L., Dümmler, B. A., Jensen, C. J., Deak, M., Gammeltoft, S., Biondi, R. M., *EMBO J.* **2002**, *21*, 5396-5407.

<sup>22</sup> Hu, Y., Furtmann, N. and Bajorath, J., *J. Med. Chem.* **2015**, *58*, 30-40.

<sup>23</sup> Cowan-Jacob, S. W., Jahnke, W., Knapp, S., *Future Med. Chem.* **2014**, *6*, 541-561.

<sup>24</sup> Singh, J., Petter, R. C., Baillie, T. A., Whitty, A., *Nat. Rev. Drug Discov.* **2011**, *10*, 307-317.

<sup>25</sup> Wood, E. R., Truesdale, A. T., McDonald, O. B., Yuan, D., Hassell, A., Dickerson, S. H., Ellis, B., Pennisi, C., Horne, E., Lackey, K., Alligood, K. J., Rusnak, D. W., Gilmer, T. M., Shewchuk, L., *Cancer Res.* **2004**, *64*, 6652-6659.

<sup>26</sup> Keen N. and Taylor S., *Nat. Rev. Cancer* **2004**, *4*, 927-936.

<sup>27</sup> Yan, M., Wang, C., He, B., Yang, M., Tong, M., Long, Z., Liu, B., Peng, F., Xu, L., Zhang, Y., Liang, D., Lei, H., Subrata, S., Kelley, K. W., Lam, E. W. F., Jiu, B., Liu, Q., *Med. Res. Rev.* **2016**, *0*, 1-44.

<sup>28</sup> Bischoff, J. R., Anderson, L., Zhu, Y. F., Mossie, K., Ng, L., Souza, B., Schryver, B., Flanagan, P., Clairvoyant, F., Ginther, C., Chan, C. S. M., Novotny, M., Slamon, D. J., Plowman, G. D., *EMBO J.* **1998**, *17*, 3052-3065.

<sup>29</sup> Dodson, C. A., Kosmopoulou, M., Richards, M. W., Atrash, B., Bavetsias, V., Blagg, J., Bayliss, R., *Biol. Chem. J.* **2010**, *427*, 19-28.

<sup>30</sup> Carmena, M. and Earnshaw, W. C., *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, 842-854.

<sup>31</sup> Crane, R., Gadea, B., Littlepage, L., Wu, H., Ruderman, J. V., *Biol. Cell* **2003**, *96*, 215-229.

<sup>32</sup> Lane, H. A. and Nigg, E. A., *J. Cell Biol.* **1996**, *135*, 1701-1713.

<sup>33</sup> Lee, M. J., Gergely, F., Jeffers, K., Peak-Chew, S. Y., Raff, J. W., *Nat. Cell Biol.* **2001**, *3*, 643-649.



- <sup>34</sup> Kufer, T. A., Silljé, H. H., Körner, R., Gruss, O. J., Meraldi, P., Nigg, E. A., *J. Cell Biol.* **2002**, 158, 617-623.
- <sup>35</sup> Kunitoku, N., Sasayama, T., Marumoto, T., Zhang, D., Honda, S., Kobayashi, O., Hatakeyama, K., Ushio, Y., Saya, H., Hirota, T., *Dev. Cell* **2010**, 19, 66-77.
- <sup>36</sup> a) Littlepage, L. E. and Ruderman, J. V., *Genes Dev.* **2002**, 16, 2274-2285; b) Floyd, S., Pines, J. and Lindon, C., *Curr. Biol.* **2008**, 18, 1649-1658.
- <sup>37</sup> Kinzel, D., Boldt, K., Davis, E. E., Burtscher, I., Trümbach, D., Diplas, B., Attié-Bitach, T., Wurst, W., Katsanis, N., Ueffing, M., Lickert, H., *Dev. Cell* **2010**, 19, 66-77.
- <sup>38</sup> Zhong W. and Chia W., *Curr. Opin. Neurobiol.* **2008**, 18, 4-11.
- <sup>39</sup> Wirtz-Peitz, F., Nishimura, T., Knoblich, J. A., *Cell* **2008**, 135, 161-173.
- <sup>40</sup> Wang, H., Somers, G. W., Bashirullah, A., Heberlein, U., Yu, F., Chia, W., *Genes Dev.* **2006**, 20, 3453-3463.
- <sup>41</sup> Burum-Auensen, E., De Angelis, P. M., Schjølberg, A. R., Kravik, K. L., Aure, M., Clausen, O. P. F., *J. Histochem. Cytochem.* **2007**, 55, 477-486.
- <sup>42</sup> Yang, G., Chang, B., Yang, F., Guo, X., Cai, K. Q., Xiao, X., Wang, H., Sen, S., Hung, M., Mills, G. B., Chang, S., Multani, A. S., Mercado-Uribe, I., Liu, J., *Clin. Cancer Res.* **2010**, 16, 3171-3181.
- <sup>43</sup> Yang, J., Ikezoe, T., Nishioka, C., Udaka, K., Yokoyama, A., *Int. J. Cancer* **2014**, 134, 1183-1194.
- <sup>44</sup> a) Stark, G. R. and Taylor, W. R., *Mol. Biotechnol.* **2006**, 32, 227-248; b) Krystyniak, A., Garcia-Echeverria, C., Prigent, C., Ferrari, S., *Oncogene* **2006**, 25, 338-348.
- <sup>45</sup> a) Yao, J., Min, Y., Guan, Z., Pan, C., Xia, L., Li, C., Wang, L., Long, Z., Zhao, Y., Li, M., Zheng, F., Xu, J., Lin, D., Liu, Q., *Mol. Cancer* **2009**, 8, 95; b) Saiprasada, G., Chitraa, P., Manikandanb, R., Sudhandiran, G., *Eur. J. Cancer* **2014**, 50, 2489-2507.
- <sup>46</sup> Guan, Z., Wang, X. R., Zhu, X. F., Huang, X. F., Xu, J., Wang, L. H., Wan, X. B., Long, Z. J., Liu, J. N., Feng, G. K., Huang, W., Zeng, Y. X., Chen, F. J., Liu, Q., *Cancer Res.* **2007**, 67, 10436-10444.
- <sup>47</sup> Braun, A., Dang, K., Buslig, F., Baird, M. A., Davidson, M. W., Waterman, C. M., Myers, K. A., *J. Cell Biol.* **2014**, 206, 97-112.
- <sup>48</sup> Kitzen, J. J. E. M., de Jonge, M. J. A., Verweiji, J., *Crit. Rev. Oncol. Hemat.* **2010**, 73, 99-110.
- <sup>49</sup> PDB: 3E5A. Zhao, B., Smallwood, A., Yang, J., Koretke, K., Nurse, K., Calamari, A., Kirkpatrick, R.B., Lai, Z. *Protein Sci.* **2008**, 17, 1791-1797.
- <sup>50</sup> Giet, R. and Pringent, C., *J. Cell Sci.* **1999**, 112, 3591-3601.
- <sup>51</sup> Yan, A., Wang, L., Xu, S., Xu, J., *Drug Discov. Today* **2011**, 16, 260-269.
- <sup>52</sup> Littlepage, L. E., Ruderman, J. V., *Genes Dev.* **2002**, 16, 2274-2285.
- <sup>53</sup> Eyers P. A., Erikson, E., Chen, L. G., Maller, J. L., *Curr. Biol.* **2003**, 13, 691-697.



- <sup>54</sup> Xu, X., Wang, X., Xiao, Z, Li, Y., Wang, Y., *PLOS. ONE* **2011**, *6*, 16757.
- <sup>55</sup> Chen, S. S., Chang, P. C., Cheng, Y. W., Tang, F. M., Lin, Y. S., *EMBO J.* **2002**, *21*, 4491-4499.
- <sup>56</sup> Goldenson, B. and Crispino, J. D., *Oncogene* **2015**, *34*, 537-545.
- <sup>57</sup> Cheung, C.H.A., Sarvagalla, S., Lee, J. Ying-Chieh, Huang, Yi-Chun, Coumar, M.S., *Expert Opin Ther Pat.* **2014**, *24*, 1021-1038.
- <sup>58</sup> a) Harrington, E. A., Bebbington, D., Moore, J., Rasmussen, R. K., Ajose-Adeogun, A. O., Nakayama, T., Graham, J. A., Demur, C., Hercend, T., Diu-Hercend, A., Su, M., Golec, J. M., Miller, K. M., *Nat. Med.* **2004**, *10*, 262-267; b) Patent: US200404932: Vertex Pharmaceuticals Inc., 130 Waverly Street, Cambridge, MA, US. *Processes for preparing substituted pyrimidines*, filed 18 June 2003; c) Bebbington, D., Binch, H., Charrier, J. D., Everitt, S., Fraysse, D., Golec, J., Kay, D., Knegetel, R., Mak, C., Mazzei, F., Miller, A., Mortimore, M., O'Donnell, M., Patel, S., Pierard, F., Pinder, J., Pollard, J., Ramaya, S., Robinson, D., Rutherford, A., Studley, J., Westcott, J., *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3586-3592.
- <sup>59</sup> Yang, J., Ikezoe, T., Nishioka, C., Tasaka, T., Taniguchi, A., Kuwayama, Y., Komatsu, N., Bandobashi, K., Togitani, K., Koeffler, P. H., Taguchi, H., and Yokoyama, A., *Blood* **2007**, *110*, 2034-2040.
- <sup>60</sup> Pollard, J. R. and Mortimore, M., *J. Med. Chem.* **2009**, *52*, 2629-2651.
- <sup>61</sup> a) Fletcher, G. C., Brox, R. D., Denny, T. A., Hembrough, T. A., Plum, S. M., Fogler, W. E., Sidor, C. F., Bray, M. R. *Mol Cancer Ther* **2011**, 126-137; b) *Targeted therapy for acute myeloid leukemia* (ed. Andreeff, M.), **2015**, Springer, New York, p.397.
- <sup>62</sup> Howard, S., Berdini, V., Boulstridge, J. A., Carr, M. G., Cross, D. M., Curry, J., Devine, L. A., Early, T. R., Fazal, L., Gill, A. L., Heathcote, M., Maman, S., Matthews, J. E., McMenamin, R. L., Navarro, E. F., O'Brien, M. A., O'Reilly, M., Rees, D. C., Reule, M., Tisi, D., Williams, G., Vinković, M., Wyatt, P. G., *J. Med. Chem.* **2009**, *52*, 379-388.
- <sup>63</sup> Giles, F. J., Cortes, J., Jones, D., Bergstrom, D., Kartajian, H., Freedman, S. J., *Blood* **2007**, *109*, 500-502.
- <sup>64</sup> Manfredi, M. G., Ecsedy, J. A., Meetze, K. A., Balani, S. K., Burenkova, O., Chen, W., Galvin, K. M., Hoar, K. M., Huck, J. J., LeRoy, P. J., Ray, E. T., Sells, T. B., Stringer, B., Stroud, S. G., Vos, T. J., Weatherhead, G. S., Wysong, D. R., Zhang, M., Bolen, J. B., Claiborne, C. F., *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 4106-4111.
- <sup>65</sup> a) Sells, T. B., Chau, R., Ecsedy, J. A., Gershman, R. E., Hoar, K., Huck, J., Janowick, D. A., Kadambi, Vivek J., LeRoy, P. J., Stirling, M., Stroud, S. G., Vos, T. J., Weatherhead, G. S., Wysong, D. R., Zhang, M., Balani, S. K., Bolen, J. B., Manfredi, M. G., Claiborne, C. F., *ACS Med. Chem. Lett.* **2015**, *6*, 630-634; b) Clinical trial numbers: NCT00652158 and NCT00249301 (ClinicalTrials.gov identifier).

- <sup>66</sup> Manfredi, M. G., Ecsedy, J. A., Chakravarty A, Silverman, L., Zhang, M., Hoar, K. M., Stroud, S. G., Chen, W., Shinde, V., Huck, J. J., Wysong, D. R., Janowick, D. A., Hyer, M. L., Leroy, P. J., Gershman, R. E., Silva, M. D., Germanos, M. S., Bolen, J. B., Claiborne, C. F., Sells, T. B., *Clin. Cancer Res.* **2011**, *17*, 7614-7624.
- <sup>67</sup> Matulonis, U. A., Sharma, S., Ghamande, S., Gordon, M. S., del Prete, S. A., Ray-Coquard, I., Kutarska, E., Liu, H., Fingert, H., Zhou, X., Danaee, H., Schilder, R. J., *Gynecol. Oncol.* **2012**, *127*, 63-69.
- <sup>68</sup> Fletcher, G. C., Brokx, R. D., Denny, T. A., Hembrough, T. A., Plum, S. M., Fogler, W. E., Sidor, C. F., Bray, M. R., *Mol. Cancer Ther.* **2011**, *10*, 126-137.
- <sup>69</sup> Dennis, M., Davies, M., Oliver, S., D'Souza, R., Pike, L., Stockman, P., *Cancer Chemother. Pharmacol.* **2012**, *70*, 461-469.
- <sup>70</sup> a) Walsby, E., Walsh, V., Pepper, C., Burnett, A., Mills, K., *Haematologica.* **2008**, *93*, 662-669.; b) Hartsink-Segers, S. A., Zwaan, C. M., Exalto, C., Luijendijk, M. W., Calvert, V. S., Petricoin, E. F., Evans, W. E., Reinhardt, D., de Haas, V., Hedtjörn, M., Hansen, B. R., Koch, T., Caron, H. N., Pieters, R., Den Boer, M. L., *Leukemia.* **2013**, *27*, 560-568.
- <sup>71</sup> a) Tsuboi, K., Yokozawa, T., Sakura, T., Watanabe, T., Fujisawa, S., Yamauchi, T., Uike, N., Ando, K., Kihara, R., Tobina, K., Asou, H., Hotta, T., Miyawaki, S., *Leuk. Res.* **2011**, *35*, 1384-1389; b) Löwenberg, B., Muus, P., Ossenkoppele, G., Rousselot, P., Cahn, J. Y., Ifrah, N., Martinelli, G., Amadori, S., Berman, E., Sonneveld, P., Jongen-Lavrencic, M., Rigaudeau, S., Stockman, P., Goudie, A., Faderl, S., Jabbour, E., Kantarjian, H., *Blood* **2011**, *118*, 6030-6036.
- <sup>72</sup> a) Sloane, D. A., Trikić, M. Z., Chu, M. L., Lamers, M. B., Mason, C. S., Mueller, I., Savory, W. J., Williams, D. H.; Eysers, P. A., *ACS Chem. Biol.* **2010**, *18*, 563-76; b) Golec, J. M. C., Discovery of MK-0457 (VX680), in *Kinase Inhibitor Drugs* (eds Li, R. and Stafford, J. A.), **2009**, John Wiley & Sons, Inc., Hoboken, NJ, USA.
- <sup>73</sup> Unpublished experimental data from Dr. Tsuchiya.
- <sup>74</sup> Results obtained kinase profiling screen performed at the International Centre for Kinase Profiling, University of Dundee; provided by Dr. Tsuchiya.
- <sup>75</sup> Leeson, P., *Nature* **2012**, *481*, 455-456.
- <sup>76</sup> *Introduction to bioinformatics 3<sup>rd</sup> Edition*, Lesk, A. M., **2008**, Oxford University Press, Great Clarendon Street, Oxford, UK, OX2 6DP.
- <sup>77</sup> *Lehninger Principles of Biochemistry 6<sup>th</sup> Edition*, Melson, D. L. and Cox, M. M., **2013**, Mcmillan Higher Education, Houndmills, Basingtoke, England, RG21 6XS.
- <sup>78</sup> a) Pauling L. and Corey R. B., *Proc. Natl. Acad. Sci. USA* **1951**, *37*, 729-740; b) Pauling L., Corey, R. B. and Branson H. R., *Proc. Natl. Acad. Sci. USA* **1951**, *37*, 205-211.
- <sup>79</sup> Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V., *J. Mol. Biol.* **1963**, *7*, 95-99.
- <sup>80</sup> Tsai, C. J., Norel, R., Wolfson, H. J., Maizel, J. V., Nussinov, R. **2001**. Protein-Ligand Interactions: Energetic Contributions and Shape Complementarity. *eLS*.

- <sup>81</sup> Tsai, C. J., Lin, S. L., Wolfson, H. and Nussinov, R. *Protein Sci.* **1997**, *6*, 51-62.
- <sup>82</sup> Hubbard, R. E. and Kamran Haider, M., **2010**. Hydrogen Bonds in Proteins: Role and Strength. *eLS*.
- <sup>83</sup> Liu, Z., Wang, G., Li, Z., Wang, R., *J. Chem. Theory Comput.* **2008**, *4*, 1959-1973.
- <sup>84</sup> Fleming, P. J. and Rose, G. D., *Protein Science* **2005**, *14*, 1911-1917.
- <sup>85</sup> a) Kleywegt, G. J., and Jones, T. A., *Methods Enzymol.* **1997**, *277*, 208-230; b) Kleywegt, G., *Practical "Model Validation" - EMBO Bioinformatics Course - Uppsala* **2001** - © 2001-2014.
- <sup>86</sup> Morris, A. L., MacArthur M. W., Hutchinson, E. G., Thornton, J. M., *Proteins* **1992**, *12*, 345-364.
- <sup>87</sup> Anderson, A. C., *Chem. Biol.* **2003**, *10*, 787-797.
- <sup>88</sup> Kleywegt, G. J., Bergfors, T., Senn, H., Le Motte, P., Gsell, B., Shudo, K., Jones, T. A., *Structure* **1994**, *2*, 1241-1258.
- <sup>89</sup> Brünger, A. T. and Rice, L. M., *Methods Enzymol.* **1997**, *277*, 243-269.
- <sup>90</sup> Pearson W. R. and Lipman D. J. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 2444-2448.
- <sup>91</sup> Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J., *J. Mol. Biol.* **1990**, *215*, 403-410.
- <sup>92</sup> Thompson, J. D., Higgins, D. G. and Gibson, T. J., *Nucleic Acids Res.* **1994**, *22*, 4673-4680.
- <sup>93</sup> PDB: 4FR4. Chaikuad, A., Elkins, J. M., Krojer, T., Mahajan, P., Goubin, S., Szklarz, M., Tumber, A., Wang, J., Savitsky, P., Shrestha, B., Daga, N., Picaud, S., Fedorov, O., Allerston, C.K., Latwiel, S. V. A., Vollmar, M., Canning, P., von Delft, F., Arrowsmith, C. H., Edwards, A. M., Bountra, C., Knapp, S., *Structural Genomics Consortium* (to be published).
- <sup>94</sup> a) Cole, J. C., Nissink, J. W. M. and Taylor, R. in *Virtual Screening in Drug Discovery* (Eds. B. Shoichet, J. Alvarez), **2005**, Taylor & Francis CRC Press, Boca Raton, Florida, USA; b) Jones, G., Willett, P. and Glen, R. C., *J. Mol. Biol.* **1995**, *245*, 43-53; c) Jones, G., Willett, P., Glen, R. C., Leach, A. R. and Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727-748; d) Nissink, J. W. M., Murray, C., Hartshorn, M., Verdonk, M. L., Cole, J. C., Taylor, R., *Proteins* **2002**, *49*, 457-471; e) Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., Taylor, R. D., *Proteins* **2003**, *52*, 609-623; f) Verdonk, M. L., Chessari, G., Cole, J. C., Hartshorn, M. J., Murray, C. W., Nissink, J. W. M., Taylor, R. D., Taylor, R., *J. Med. Chem.* **2005**, *48*, 6504-6515; g) Hartshorn, M. J., Verdonk, M. L., Chessari, G., Brewerton, S. C., Mooij, W. T. M., Mortenson, P. N., Murray, C. W., *J. Med. Chem.* **2007**, *50*, 726-741.
- <sup>95</sup> Nissink, J. W. M., Murray, C., Hartshorn, M., Verdonk, M. L., Cole, J. C., Taylor, R., *Proteins* **2002**, *49*, 457-471.
- <sup>96</sup> Hartshorn, M. J., Verdonk, M. L., Chessari, G., Brewerton, S. C., Mooij, W. T. M., Mortenson, P. N., Murray, C. W., *J. Med. Chem.* **2007**, *50*, 726-741.
- <sup>97</sup> Kramer B., Rarey M., Lengauer T., *Proteins* **1999**, *37*, 228-241.
- <sup>98</sup> Korb, O., Stützel, T., and Exner, T. E. *J. Chem. Inf. Model.* **2009**, *49*, 84-96.

- <sup>99</sup> PDB 3EFW. Cee, V. J., Cheng, A. C., Romero, K., Bellon, S., Mohr, C., Whittington, D. A., Bak, A., Bready, J., Caenepeel, S., Coxon, A., Deak, H. L., Fretland, J., Gu, Y., Hodous, B. L., Huang, X., Kim, J. L., Lin, J., Long, A. M., Nguyen, H., Olivieri, P. R., Patel, V. F., Wang, L., Zhou, Y., Hughes, P., Geuns-Meyer, S., *Bioorg. Med. Chem. Lett.* **2009**, *19*, 424-427.
- <sup>100</sup> PDB: 4JBP. Wu, J. M., Chen, C. T., Coumar, M. S., Lin, W. H., Chen, Z. J., Hsu, J. T., Peng, Y. H., Shiao, H. Y., Lin, W. H., Chu, C. Y., Wu, J. S., Lin, C. T., Chen, C. P., Hsueh, C. C., Chang, K. Y., Kao, L. P., Huang, C. Y., Chao, Y. S., Wu, S. Y., Hsieh, H. P., Chi, Y. H., *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 1779-1787.
- <sup>101</sup> Laskowski, R. A., MacArthur, M. W., Moss, D. S., Thornton, J. M., *J. Appl. Crystal.* **1993**, *26*, 283-291.
- <sup>102</sup> *Molecular Operating Environment (MOE)*, 2013.08; **2016** Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7.
- <sup>103</sup> Laskowski, R. A., Swindells, M., *J. Chem. Inf. Model.* **2011**, *51*, 2778-2786
- <sup>104</sup> a) Gaulton, A., Bellis, L., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Akhtar, R., Bento, A.P., Al-Lazikani, B., Michalovich, D., Overington J.P., *Nucleic Acids Res. Database Issue* **2012**, *40*, 1100-1107; b) Bento, A.P., Gaulton, A., Hersey, A., Bellis, L.J., Chambers, J., Davies, M., Krüger, F.A., Light, Y., Mak, L., McGlinchey, S., Nowotka, M., Papadatos, G., Santos, R., Overington, J.P., *Nucleic Acids Res. Database Issue* **2014**, *42*, 1083-1090.
- <sup>105</sup> de Beer, T. A. P., Berka, K., Thornton, J. M., Laskowski, R. A., *Nucleic Acids Res.* **2014**, *42*, 292-296.
- <sup>106</sup> Laskowski, R. A., *Nucleic Acids Res.* **2001**, *29*, 221-222.
- <sup>107</sup> www.rcsbpdb.org Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., *The Protein Data Bank* **2000**.
- <sup>108</sup> PerkinElmer Chemdraw Professional and ChemOffice 2010.
- <sup>109</sup> Viswanadhan, V. N., Ghose, A. K., Revankar, G. R. and Robins, R. K., *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163-172.
- <sup>110</sup> PDB: 4BN1. Rowan, F.C., Richards, M., Bibby, R.A., Thompson, A., Bayliss, R., Blagg, J., *ACS Chem. Biol.* **2013**, *8*, 2184-2191.
- <sup>111</sup> PDB: 4C2V and 4C2W. Sessa, F., Villa, F., *Acta Crystallogr. Sect. F* **2014**, *70*, 294-298.
- <sup>112</sup> PDB: 2H8H. Hennequin, L. F., Allen, J., Breed, J., Curwen, J., Fennell, M., Green, T. P., Lambert-van der Brempt, C., Morgentin, R., Norman, R. A., Olivier, A., Otterbein, L., Ple, P. A., Warin, N., Costello, G., *J. Med. Chem.* **2006**, *49*, 6465-6488.
- <sup>113</sup> PDB: 5AM6. Bunney, T., Wan, S., Thiyagarajan, N., Sutto, L., Williams, S. V., Ashford, P., Koss, H., Knowles, M. A., Gervasio, F. L., Coveney, P. V., Katan, M., *Ebiomedicine*, **2015**, *2*, 194-204.

- <sup>114</sup> PDB: 4I3Z. Jacobsen, D. M., Bao, Z. Q., O'Brien, P. J., Brooks, C. L., Young, M. A. *J. Am. Chem. Soc.* **2012**, *134*, 15357-15370.
- <sup>115</sup> PDB: 2GU8. Lin, X., Murray, J. M., Rico, A. C., Wang, M. X., Chu, D. T., Zhou, Y., Del Rosario, M., Kaufman, S., Ma, S., Fang, E., Crawford, K., Jefferson, A. B., *Bioorg. Med. Chem. Lett.* **2006**, *16*, 4163-4168.
- <sup>116</sup> a) *Bioconjugate Techniques, 3rd Edition* by Greg T. Hermanson, **2013**, Academic Press, 32 Jamestown Road, London NW1 7BY, UK; b) Lau, E. P., Haley, B. E., Barden, R. E., *Biochemistry* **1977**, *16*, 2581-2585.
- <sup>117</sup> PDB: 2DWB. Kukimoto-Niino, M., Murayama, K., Shirouzu, M., Yokoyama, S. "Aurora-A kinase complexed with AMPPNP" (to be published).
- <sup>118</sup> a) Erixon, K. M., Dabalos, C. L., Leeper, F. J., *Chem. Commun.* **2007**, *9*, 960-962; b) Swarbrick, J. M., Graeff, R., Garnham, C., Thomas, M. P., Galione, A., Potter, B. V. L., *Chem. Commun.* **2014**, *50*, 2458-2461; c) El-Sagheer, A. H., Brown, T., *Q. Rev. Biophys.* **2015**, *48*, 429-436; d) Mustiya, D., Selvam, C., Kennedy, S.D., Rozners, E., *Bioorg. Med. Chem. Lett.* **2011**, *21*, 3420-3422.
- <sup>119</sup> Dr. Edith Chan's in-house database for kinase inhibitor scaffolds.
- <sup>120</sup> Virga, K. G., Zhang, Y. M., Leonardi, R., Ivey, R. A., Hevener, K., Park, H. W., Jackowski, S., Rock C. O. and Lee, R. E., *Bioorgan. Med. Chem.* **2006**, *14*, 1007-1020.
- <sup>121</sup> Kopelevich, V. M., Evdokimova, G. S., Marieva, T. D., Shmuilovich, L. M., *Pharm. Chem. J.* **1971**, *5*, 534; b) Hoeper, F., Montforts, F. P., *Tetrahedron Asymm.* **1993**, *4*, 1439-1440.
- <sup>122</sup> Joullie, M. M. and Lassen, K. M., *Evolution of Amide Bond Formation*. ARKIVOC, **2010**, *8*, 189-250.
- <sup>123</sup> Shiori, T., Ninomiya, K., Yamada, S., *J. Am. Chem. Soc.* **1972**, *94*, 6203.
- <sup>124</sup> Belleau, B., Malek, G., *J. Am. Chem. Soc.* **1968**, *90*, 1651.
- <sup>125</sup> Knorr, R., Trzeciak, A., Bannwarth, W., Gillessen, D., *Tetrahedron Lett.* **1989**, *30*, 1927-1930.
- <sup>126</sup> Carpino, L.A., *J. Am. Chem. Soc.* **1993**, *115*, 4397-4398.
- <sup>127</sup> Sheehan, J. C. and Hess, G. P., *J. Am. Chem. Soc.* **1955**, *77*, 1067-1068; b) Anderson, G. W. and Callahan, F. M., *J. Am. Chem. Soc.* **1958**, *80*, 2902; c) Schön, I. and Kisfaludy, L., *Synthesis* **1986**, *4*, 303-305.
- <sup>128</sup> a) Seebach, D., Abele, S., Sifferlen, T., Hanggi, M., Gruner, S., Seiler, P., *Helv. Chim. Acta* **1998**, *81*, 2218-2243; b) Bavikar, S. N., Salunke, D. B., Hazra, B. G., Pore, V. S., Dodd, R. H., J., Shirazi, F., Deshpande, M. V., Kadreppa, S., Chattopadhyay, S., *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5512-5517.
- <sup>129</sup> Gopinathan, M. B., Jin, C., Rehder, K. S., *Synth. Commun.* **2009**, *38*, 3973-3981.
- <sup>130</sup> Rubin, S. H., *J. Pharm. Sci.* **1948**, *37*, 502-504.

- <sup>131</sup> a) Hartwig, J. F., *Angew. Chem. Int. Ed.* **1998**, *37*, 2046-2067; b) Hartwig, J. F., Kawatsura, M., Hauck, S. I., Shaughnessy, K. H., Alcazar-Roman, L. M., *J. Org. Chem.* **1999**, *64*, 5575-5580; c) Wolfe, J. P. and Buchwald, S. L., *J. Org. Chem.* **2000**, *65*, 1144-1157; d) Wolfe, J. P., Tomori, H., Sadighi, J. P., Yin, J., Buchwald, S. L., *J. Org. Chem.* **2000**, *65*, 1158-1174; e) Muci, A. R., Buchwald, S. L., *Topics in Curr. Chem.* **2002**, *219*, 131-209.
- <sup>132</sup> Delia, T. J., Stark, D., Glenn, S. K., *J. Heterocycl. Chem.* **1995**, *32*, 1177-1180.
- <sup>133</sup> a) Paul, F., Patt, J., Hartwig, J. F., *J. Am. Chem. Soc.* **1994**, *116*, 5969-5970; b) Guram, A. S., Rennels, R. A., Buchwald, S. L., *Angew. Chem. Int. Ed.* **1995**, *34*, 1348-1350; c) Louie, J., Hartwig, J. F., *Tetrahedron Lett.* **1995**, *36*, 3609-3612.
- <sup>134</sup> Beare, N. A., Hartwig, J. F., *J. Org. Chem.* **2002**, *67*, 541-555.
- <sup>135</sup> Wolfe, J. P., Wagaw, S., Buchwald, S. L., *J. Am. Chem. Soc.* **1996**, *118*, 7215-7216.
- <sup>136</sup> Singh, U. K., Strieter, E. R., Blackmond, D. G., Buchwald, S. L., *J. Am. Chem. Soc.* **2002**, *124*, 14104-14114.
- <sup>137</sup> Moss, T. A., Addie, M. S., Nowak, T., Waring, M. J., *Synlett* **2012**, *2*, 285-289.
- <sup>138</sup> Sagi, V. V., Liu, T., Lu, X., Bartfai, T., Roberts, E., *Bioorg. Med. Chem. Lett.* **2011**, *21*, 7210-7215.
- <sup>139</sup> Ikawa, T., Barder, T. E., Biscoe, M. R., Buchwald, S. L., *J. Am. Chem. Soc.* **2007**, *129*, 13001-13007.
- <sup>140</sup> Murray, P. M., Bellany, Bellany, F., Benhamou, L., Bucar, D. K., Tabor, A. B., Sheppard, T. D., *Org. Biomol. Chem.* **2016**, *14*, 2373-2384.
- <sup>141</sup> White, J. D., Hansen, J. D., *J. Am. Chem. Soc.* **2002**, *124*, 4950-4951.
- <sup>142</sup> Mittapalli, G. K., Osornio, Y. M., Guerrero, M. A., Reddy, K. R., Krishnamurthy, R., Eschenmoser, A., *Angew. Chem. Int. Ed.* **2007**, *46*, 2470-2477.
- <sup>143</sup> *Organic Chemistry – Breakthroughs and Perspectives, 1<sup>st</sup> Edition.* (Eds. Kuiling Ding and Li-Xin Dai), **2012**, Wiley-VCH, UK. *Chapter 7: “CuAAC: the Quintessential Click Reaction”*, Fokin, V.V.
- <sup>144</sup> Rostovtsev, V. V., Green, L. G., Fokin, V. V., Sharpless, K.B., *Angew. Chem. Int. Ed.* **2002**, *41*, 2596-2599.
- <sup>145</sup> Appukkuttan, P., Dehaen, W., Fokin, V. V., and van der Eycken, E. *Org. Lett.* **2004**, *6*, 4223-4225.
- <sup>146</sup> a) Chan, T. R., Hilgraf, R., Sharpless, K. B., Fokin, V. V., *Org. Lett.* **2004**, *6*, 2853-2855; b) Hein, J. E., Krasnova, L. B., Iwasaki, M., Fokin, V.V., *Org. Synth.* **2011**, *88*, 238-246.
- <sup>147</sup> Choi, I., Kim, Y. K., Min, D. H., Lee, S., Yeo, W. S., *J. Am. Chem. Soc.* **2011**, *133*, 16718-16721.
- <sup>148</sup> Ishikawa, F., Kakeya, H., *Bioorg. Med. Chem. Lett.* **2014**, *24*, 865-869.
- <sup>149</sup> Jawalekar, A. M., Meeuwenoord, N., Cremers, J. G. O., Overkleeft, H. S., van der Marel, G. A., Rutjes, F. P. J. T., van Delft, F. L., *J. Org. Chem.* **2008**, *73*, 287-290.

- <sup>150</sup> Schmieder-Van De Vondervoort L., Bouttemy, S., Padron, J.M., Le Bras, J., Muzart, J., Alsters, P.L., *Synlett* **2002**, 2, 243.
- <sup>151</sup> Patent US3210405A: Lee M. A., Dayton, O., assignor to Monsanto Company, a corporation of Delaware. *Unsaturated esters of propiolic acid*, filed Jan. 3, 1961.
- <sup>152</sup> Yi, L., Abootorabi, M., Wu, Y., *Chem. Bio. Chem.* **2011**, 12, 2413-2417.
- <sup>153</sup> Himo, F., Lovell, T., Hilgraf, R., Rostovtsev, V. V., Noodleman, L., Sharpless, K. B., Fokin, V. V., *J. Am. Chem. Soc.* **2005**, 127, 210-216.
- <sup>154</sup> a) Mustiya, D., Selvam, C., Kennedy, S. D., Rozners, E., *Bioorg. Med. Chem. Lett.* **2011**, 21, 3420-3422; b) Mathew, S. C., By, Y., Berthault, A., Virolleaud, M., Carrega, L., Chouraqui, G., Commeiras, L., Condo, J., Attolini, M., Gaudel-Siri, A., Ruf, J., Rodriguez, J., Parrain, J., Guieu, R., *Org. Biomol. Chem.* **2010**, 8, 3874-3881; c) Dyrager, C., Borjesson, K., Diner, P., Elf, A., Albinsson, B., Wilhelmsson, L. M., Grotli, M., *Eur. J. Org. Chem.* **2009**, 10, 1515-1521; d) Hou, S., Liu, W., Ji, D., Zhao, Z., *Bioorg. Med. Chem. Lett.* **2011**, 21, 1667-1669.
- <sup>155</sup> O' Mahony, G., Svensson, S., Sundgren, A., Grotli, M., *Nucleos. Nucleot. Nucl.* **2008**, 27, 449-459.
- <sup>156</sup> Tam, A., Arnold, U., Soellner, M. B., Raines, R. T., *J. Am. Chem. Soc.* **2007**, 129, 12670-12671.
- <sup>157</sup> Solařová, H., Císařová, I., Štěpnička, P., *Organometallics* **2014**, 33, 4131-4147.
- <sup>158</sup> Mistry, S. N., Baker, J. G., Fischer, P. M., Hill, S.J., Gardiner, S. M., Kellam, B., *J. Med. Chem.* **2013**, 56, 3852-3865.
- <sup>159</sup> *Handbook of Reagents for Organic Synthesis, Reagents for Silicon-mediated Organic Synthesis* (Fuchs, Philip L. editor), **2011**, Wiley, ISBN: 978-0-470-71023-4, p.232.
- <sup>160</sup> Stiller, E. T., Harris, S. A., Finkelstein, J., Keresztesy, J. C., Folkers, K., *J. Am. Chem. Soc.* **1940**, 62, 1785-1790.
- <sup>161</sup> Fujino, H., *Chem. Pharm. Bull.* **1967**, 15, 2015-2016.
- <sup>162</sup> Marastoni, M., Guerrini, R., Balboni, G., Salvadori, S., Fantin, G., Fogagnolo, M., Lazarus, L. H., Tomatis, R., *Arzneimittel-Forschung/Drug Research* **1999**, 49, 6-12.
- <sup>163</sup> Casiraghi, G., Rassa, G., Auzzas, L., Burreddu, P., Gaetani, E., Battistini, L., Zanardi, F., Curti, C., Nicastro, G., Belvisi, L., Motto, I., Castorina, M., Giannini, G., Pisano, C., *J. Med. Chem.* **2005**, 48, 7675-7687.
- <sup>164</sup> Patent: WO2008/135786 A1: AstraZeneca AB; AstraZeneca UK Limited. *Amino-thiazolyl-pyrimidin derivatives and their use for the treatment of cancer*, filed May 2, 2008.
- <sup>165</sup> van der Does, L., den Hertog, H. J., *Tetrahedron Lett.* **1972**, 13, 2183-2186.
- <sup>166</sup> Belshawa, P. J., Mzengezaa, S., and Lajoiea, G. A., *Synth. Commun.* **1990**, 20, 3157-3160.
- <sup>167</sup> Forster, M. O., Fierz, H. E., *J. Chem. Soc.* **1908**, 93, 72-85.
- <sup>168</sup> Patent: US2482069: Ruskin, S. L., assignor to Ruskin, F.R., New York. *Preparation of adenylic and thio-adenylic acid*, filed Nov. 6, 1944.

- <sup>169</sup> Sako, M., Saito, T., Kameyama, K., Hirota, K., Maki, Y., *J. Chem. Soc., Chem. Commun.* **1987**, *17*, 1298-1299.
- <sup>170</sup> Chen, L. Q., Wilson, D. J., Xu, Y. L., Aldrich, C. C., Felczak, K., Sham, Y. Y., Pankiewicz, K. W., *J. Med. Chem.* **2010**, *53*, 4768-4778.
- <sup>171</sup> Patent: EP2471804 A1: Institut Pasteur, 75015 Paris (FR), Centre National de la Recherche Scientifique 75016 Paris (FR), Institute Curie 75248 Paris Cedex 05 (FR). *Diadenosine antibacterial compounds*, filed Jul. 4, 2012.
- <sup>172</sup> Barker, P. L., Gendler, P. L., Rapoport, H., *J. Org. Chem.* **1981**, *46*, 2455-2465.
- <sup>173</sup> Matingly, P. G., *Synthesis* **1990**, *4*, 366-368.
- <sup>174</sup> Wang, J., Uttamchandani, M., Li, J., Hu, M., Yao, S. Q., *Chem. Commun.* **2006**, *36*, 3783-3785.
- <sup>175</sup> Kopka, K., Wagner, S., Riemann, B., Law, M. P., Puke, C., Luthra, S. K., Pike, V. W., Wichter, T., Schmitz, W., Schober, O., Schaefer, M., *Bioorg. Med. Chem.* **2003**, *11*, 3513-3527.
- <sup>176</sup> Teramura, K., *Kygyo Kagaku Zasshi* **1954**, *57*, 405.
- <sup>177</sup> Castaneda, L., Wright, Z. V. F., Marculescu, C., Tran, T. M., Chudasama, V., Maruani, A., Hull, E. A., Nunes, J. P. M., Fitzmaurice, R. J., Smith, M. E. B., Jones, L. H., Caddick, S., Baker, J. R., *Tetrahedron Lett.* **2013**, *54*, 3493-3495.
- <sup>178</sup> a) Larsen, A. F. and Ulven, T., *Chem. Commun.* **2014**, *50*, 4997; b) Zhang, L., Fan, J. H., Vu, K., Hong, K., Le Brazidec, J. Y., Shi, J. D., Biamonte, M., Busch, D. J., Lough, R. E., Grecko, R., Ran, Y. Q., Sensintaffar, J. L., Kamal, A., Lundgren, K., Burrows, F. J., Mansfield, R., Timony, G. A., Ulm, E. H., Kasibhatla, S. R., Boehm, M. F., *J. Med. Chem.* **2006**, *49*, 5352-5362.