# ON THE ORIGINS OF ENZYME INHIBITOR SELECTIVITY AND PROMISCUITY: A CASE STUDY OF PROTEIN KINASE BINDING TO STAUROSPORINE

This dissertation is submitted for the degree of
Doctor of Philosophy

by

Duangrudee Tanramluk

UNIVERSITY OF CAMBRIDGE

Hughes Hall
Cambridge - England
August 2009

# DECLARATION

This dissertation is a summary of research carried out in the Department of Biochemistry, University of Cambridge, between October 2005 and January 2009. It is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. It has not, either in part, or as a whole, been submitted for a degree, diploma, or other qualification at any other university. The length of this dissertation does not exceed the word limit.

Duangrudee Tanramluk

Cambridge, England

August 20, 2009

# ABSTRACT

Protein kinases are important regulatory enzymes in signal transduction and in cell regulation. Understanding inhibition mechanisms of kinases is important for the further development of new therapies for cancer and inflammatory diseases. I have developed a statistical approach based on the Mantel test to find the relationship between the shapes of ATP binding sites and their affinities for inhibitors. My shape-based dendrogram shows clustering of the kinases based on similarity in shape. I investigate the pocket in terms of conservation of surrounding amino acids and atoms in order to identify the key determinants of ligand binding. I find that the most conserved regions are the main chain atoms in the hinge region and I show that the tetrahydropyran ring of staurosporine causes induced-fit of the glycine rich loop. I apply multiple linear regression to select distances measured between the distinctive parts of residues which correlate with the binding constants. This method allows me to understand the importance of the size of the gatekeeper residue and the closure between the first glycine of the GXGXXG motif and the aspartate of the DFG loop, which act together to promote tight binding to staurosporine. I also find that the greater the number of hydrogen bonds made by the kinase around the methylamine group of staurosporine, the tighter the binding to staurosporine. The website I have developed allows a better understanding of cross reactivity and may be useful for narrowing down the options for a synthetic strategy to design kinase inhibitors.

# ACKNOWLEDGMENTS

I wish to express my sincere gratitude to Professor Sir Tom Blundell for giving me the opportunity to study my PhD at the University of Cambridge. This thesis would not be possible without his generous support and invaluable guidance through all aspects of the research. He has been an excellent mentor and a role model for me.

I am grateful to Dr. William Pitt and Dr. Ben Luisi who have critically read my thesis chapter and made important suggestions.

I would like to acknowledge Professor Amiram Goldblum for his contribution in Hansch equation.

I would like to thank Dr. David Burke, Dr. Richard Smith, and Dr. Rinaldo Montalvao for the development of supporting programs.

I would like to express my sincere appreciation to Adrian Schreyer, Sungsam Gong, Alicia Higueruelo, Semin Lee, Juok Cho, Dr. Richard Bickerton, Dr. Alan Brown, and Anjum Karmali who widen my horizons and help me tackle the difficult part of my work. I would like to thank Mingming, Noha, Jawon, Ann, James, Qian, Takashi, and Bernardo for their cordial friendship. Special thanks go to Graham Eliff for computer support. These colleagues make my multidisciplinary research a wonderful working experience.

I would like to thank the Royal Thai Government for my PhD scholarship.

I feel particularly indebted to my father and my aunt for their ever-lasting love, understanding and encouragement. Finally, I wish to thank my sister, Sawalee, who has supported me with kindness and remarkable patience.

*To my mother, Nongluck Tanramluk, whose death by cancer inspired me to pursue a degree in drug design*

# TABLE OF CONTENTS

## LIST OF FIGURES

# LIST OF TABLES

## ABBREVIATIONS

| | |
|---|---|
| 3D | three-dimensional |
| A | alanine |
| ABL1 | v-abl Abelson murine leukemia viral oncogene homolog 1 |
| ACP | phosphomethylphosphonic acid adenylate ester |
| ADP | adenosine 5´-diphosphate |
| AGC | kinase group containing PKA, PKG, PKC as members |
| AK | atypical kinase |
| Ala | alanine |
| AMP | adenosine 5´-monophosphate |
| ANP | phosphoaminophosphonic acid-adenylate ester |
| APE | alanine-proline-glutamate |
| Arg | arginine |
| Asn | asparagine |
| Asp | aspartate |
| ATP | adenosine 5´-triphosphate |
| BUB1 | budding uninhibited by benzimidazoles 1 homolog |
| C | cysteine, carbon |
| CAMK | calcium/calmodulin-dependent protein kinase |
| CATH | class, architecture, topology and homologous superfamily |
| CDK | cyclin-dependent kinase |
| CKI | casein kinase I |
| CLK | CDC-like kinase |
| CMGC | kinases group contains CDK, MAPK, GSK3, CLK as members |
| CNS | central nervous system |
| CSK | c-src tyrosine kinase (c-Src) |
| CT | sp$^3$ carbon |
| Cys | cysteine |

| | |
|---|---|
| Cα | alpha carbon |
| D | aspartate |
| DAPK | death-associated protein kinase |
| DFG | aspartate-phenylalanine-glycine |
| DNA | deoxyribonucleic acid |
| $D_{X\_Y}$ | distance between residues X and Y |
| E | glutamic acid |
| EGFR | epidermal growth factor receptor |
| EST | expressed sequence tag |
| F | phenylalanine |
| FGFR | fibroblast growth factor receptor |
| FYN | FYN oncogene related to SRC, FGR, YES |
| G | glycine |
| Gln | glutamine |
| Glu | glutamate |
| Gly | glycine |
| GO | gene ontology |
| GSK3 | glycogen synthase kinase 3 |
| H | histidine, hydrogen |
| His | histidine |
| HIV | human immunodeficiency virus |
| I | isoleucine |
| Ile | isoleucine |
| IRK | inward rectifying K (potassium) channel |
| iTOL | interactive tree of life |
| JAK | Janus kinase |
| K | lysine |
| KAPCA | cAMP-dependent protein kinase, protein kinase A |
| $K_d$ | dissociation constant |
| $K_{d,STU}$ | dissociation constant of staurosporine in nanomolar |
| kDa | kilo-dalton |

| | |
|---|---|
| L | leucine |
| LCK | leukocyte-specific protein tyrosine kinase |
| Leu | leucine |
| log | logarithm (common, base 10) |
| Lys | lysine |
| M | methionine |
| M.W. | molecular weight |
| M3K5 | mitogen-activated protein kinase kinase kinase 5 |
| MAHORI | mapping analogous hetero-atoms onto residue interactions |
| MAPK | mitogen activated kinase |
| MEK | MAP kinase kinase or Erk Kinase |
| Met | methionine |
| MKNK2 | MAP kinase interacting serine/threonine kinase 2 |
| mol | mole |
| MSD | Macromolecular Structure Database |
| N | asparagine, nitrogen |
| $N_{4'}$ | methylamine nitrogen at the position 4′ of staurosporine |
| nM | nanomolar |
| O | oxygen |
| $O_{3'}$ | methoxy oxygen at the position 3′ of staurosporine |
| P | proline |
| PDB ID | Protein Data Bank identifier |
| Phe | phenylalanine |
| PHYLIP | phylogeny inference package |
| PI3K | phosphoinositide 3-kinases |
| PIM1 | proto-oncogene serine/threonine-protein kinase Pim-1 |
| PKA | cyclic 3'- 5' adenosine monophosphate-dependent protein kinase, or protein kinase A |
| PKC | protein kinase C |
| Pro | proline |

| | |
|---|---|
| Q | glutamine |
| QSAR | quantitative structure-activity relationship |
| R | arginine |
| RGC | receptor guanylate cyclases |
| S | serine |
| S.D. | standard deviation |
| SCOP | structural classification of proteins |
| Ser | serine |
| SH2 | Src homology 2 domain |
| SH3 | Src homology 3 domain |
| SIMFONEE | specificity implication from frequently occurring neighbouring entities |
| Src | v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog |
| STE | homologs of yeast sterile 7, 11, 20 kinases |
| STK | serine/threonine kinase |
| STU | staurosporine |
| SYK | spleen tyrosine kinase |
| T | threonine |
| T7-phage | bacteriophage T7 |
| Thr | threonine |
| TK | tyrosine kinase |
| TKL | tyrosine kinase-like |
| Trp | tryptophan |
| Tyr | tyrosine |
| V | valine |
| Val | valine |
| Y | tyrosine |

# Chapter 1

## INTRODUCTION

### 1.1 Ligand promiscuity & selectivity

Many diseases are characterised by dysregulation of biological pathways, which leads to change at the level of the individual cell, the tissue, and the whole organism. To restore the healthy state, some diseases can be cured by a small molecule drugs that inhibit a molecular target central to the disease mechanism (Zimmermann et al, 2007).

Much effort has been spent in finding a compound with high binding affinity against a single target, a magic bullet. However, for some diseases which involve several pathways such as cancer, inhibition of a single target is not enough to restore the healthy state. A recent trend has been towards designing compounds that bind to multiple defined molecular targets, a magic shotgun. The most significant effort has been on tackling the resistance problem in HIV-1 anti-viral therapy and oncology (Hopkins et al, 2006). Several failures arising from targeting a single protein and successes from many selectively non-selective drugs have led to a paradigm shift from a magic bullet to a magic shotgun approach in therapeutic intervention in cancer, cardiovascular disease and CNS disorders (Frantz, 2005; Roth et al, 2004). Although side-effects of multitarget drugs are common, it has been argued that highly cross-reactive drugs might work better against drug resistance mutations based on the fact that such drugs normally have interactions with backbone groups and evolutionary conserved residues (Zhang et al, 2008).

Protein kinases are of considerable interest to the pharmaceutical industry because dysfunction often results in malignancy (Blume-Jensen & Hunter, 2001; Cohen, 2002). This family of enzymes was chosen as my case study for the analysis of ligand promiscuity and selectivity because all share similar ATP binding sites. Indeed the success of several high affinity ATP-mimetic drugs has made the design of selective inhibitors an attractive approach to useful therapeutics, particularly for oncology (Noble et al, 2004). Although the structural conservation of the ATP binding site can lead to off-target ligand binding, kinase inhibitor design has become a promising way forward for discovery of useful therapeutic agents (Force et al, 2004).

The major challenge for protein kinase inhibitor design is obtaining selectivity. In order to reduce the chances of undesirable side effects, potency is usually optimised against a target kinase while reducing off-target activities including at other kinases. By understanding the ligand selectivity, cross reactivity might be removed by identifying a structural feature that promotes promiscuity and detecting non-conserved features that may enable paralog discrimination (Zhang et al, 2008).

## 1.2 Protein kinase

The human genome comprises more than 500 protein kinases (Manning et al, 2002), which are known to mediate most of signal transduction crucial to metabolism, cell proliferation and differentiation, membrane transport, and apoptosis. Protein kinases catalyse the transfer of a phosphate group, usually from ATP, to a hydroxyl group of a serine/threonine or tyrosine of a protein substrate. The molecular weight of protein kinase is about 30 kDa and the sequence length is usually 250-300 amino acids (Hanks et al, 1988). Protein kinases share the same fold and similar ATP binding sites. However, they exhibit a variety of conformational states and regulatory mechanisms.

### *1.2.1 Regulatory Mechanisms*

"All active kinases are alike, but an inactive kinase is inactive after its own fashion" (Noble et al, 2004). Listed below are some examples of different regulatory mechanisms in the way kinases can be controlled.

#### *1.2.1.1 By phosphorylation of the activation loop*

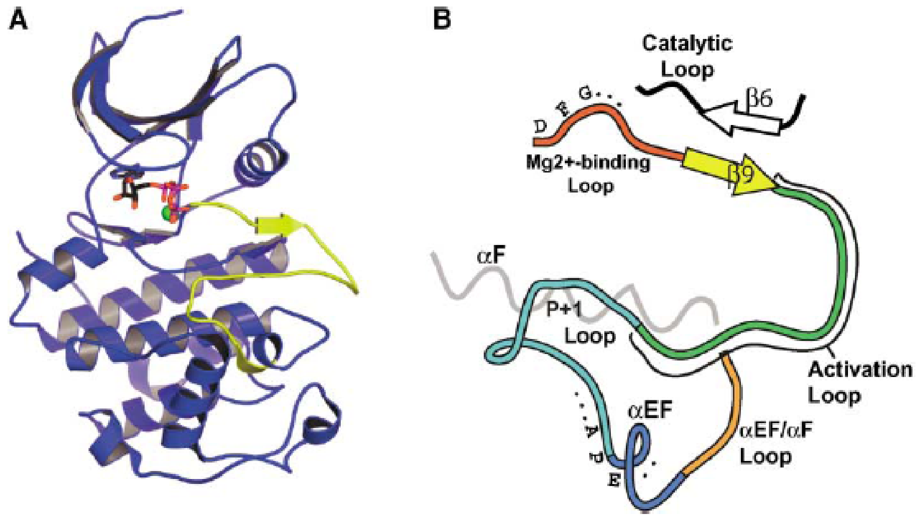


Figure 1. The yellow segment in figure 1A is the activation segment. Activation segment is defined as the region between and including the DFG and APE tripeptide motifs in figure 1B (Nolen et al, 2004).

Many kinases require activation by phosphorylation of the activation segment (Nolen et al, 2004). In the unphosphorylated state, the activation loop can adopt various conformations, for example allowing it to traverse the cleft between the N- and C-terminal lobes in IRK (Hubbard et al, 1994). Upon phosphorylation, the activation loop moves away and allows the pocket to bind both the ATP and a peptide substrate (Hubbard, 1997).

#### *1.2.1.2 Through interactions with regulatory subunits in response to second messenger*

Cyclic AMP dependent protein kinase (PKA) is one of the simplest kinases (Taylor et al, 2004) and the first protein kinase for which a crystal structure became available (Knighton et al, 1991). For these reasons, PKA are used as the reference molecule for several studies. PKA is active in the hetero-tetrameric form. Without cyclic AMP, the two regulatory subunits of PKA inhibit the two catalytic subunits by binding to their active sites. Cyclic AMP binding to the regulatory subunits causes a conformational change and releases the activated catalytic subunit.

#### *1.2.1.3 By expression level of additional subunit*

Cell cycle progression in eukaryotic systems is tightly regulated by members of the CDK family (Nurse, 2002). The unmodified CDK cannot catalyse the phosphotransfer reaction, but a CDK can be activated at particular point in cell cycle upon binding to a cyclin (Morgan, 1997).

#### *1.2.1.4 By additional domains that inhibit kinase by autoregulatory processes*

c-Src was the first proto-oncogenic tyrosine kinase to be discovered (Stehelin et al, 1976). Intramolecular interaction of SH2 inhibits Src kinase by allowing the SH3 linker to bind to the surface of the N-terminal lobe and stabilising it in an inactive conformation (Xu et al, 1999). Ligand binding to SH2/SH3 or dephosphorylation leads to the release of the activated kinase domain (Blume-Jensen & Hunter, 2001).

#### *1.2.1.5 By pseudosubstrate segment which compete with its own substrate*

The C-terminal domain of twitchin is called a pseudosubstrate because it competes directly with the substrate for access to the active site and stabilises it in an inactive conformation (Hu et al, 1994; Huse & Kuriyan, 2002).

These mechanisms demonstrate that kinases can be regulated by various types of molecules and in a variety of binding states. Biologists have attempted to categorise kinases in several different ways.

### *1.2.2 Classification*

The classic protein kinase phylogenetic tree was constructed on the basis of knowledge of sequence and biological function (Hanks et al, 1988). This thesis focuses on the relationship between the sequences and structures of kinases and inhibitor binding. Therefore, instead of classifying kinases by their function and sub-cellular location, e.g. receptor protein-tyrosine kinases (e.g. IRK, EGFR), cytoplasmic protein-tyrosine kinase (e.g. Src, JAKs, Abl), and serine-threonine kinases (e.g. CAMKs, CDK, MAPK, MEK, PKA, PKC), I refer to their type based on their sequence similarity. Kinases have been a topic of comprehensive classification (Cheek et al, 2005; Scheeff & Bourne, 2005). The most widely used human protein kinase classification phylogenetic tree is constructed based on public and proprietary genomic, complementary DNA, and expressed sequence tag (EST) sequences. (Manning et al, 2002). According to the classification of Manning, the major kinase groups are as follows.

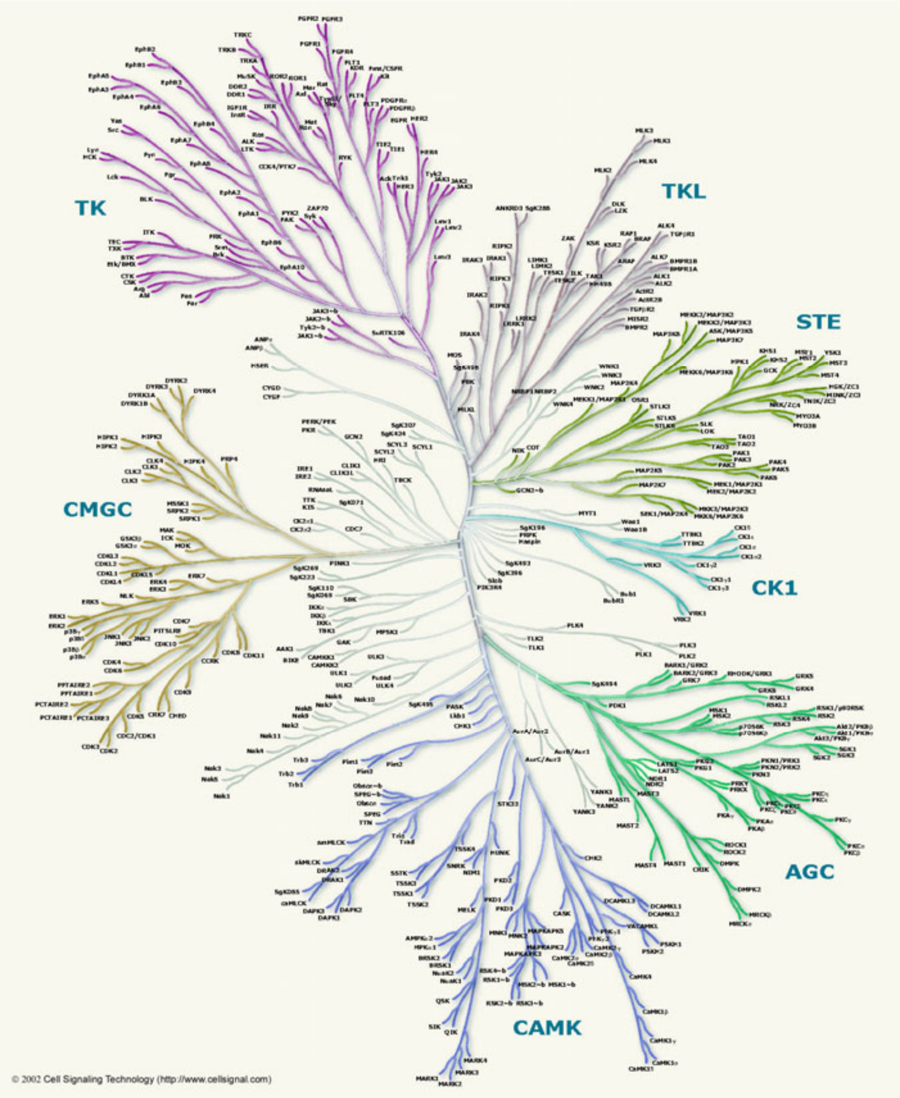

TK
TKL
STE
CMGC
CK1
AGC
CAMK
© 2002 Cell Signaling Technology (http://www.cellsignal.com)

Figure 2 The human kinase phylogenetic tree (Manning et al, 2002)

### 1.2.2.1 *TK (Tyrosine kinase)*

All members in this group phosphorylate tyrosine residues and their relationships can be observed from sequence similarity. Other dual specificity kinases (those which phosphorylate serine, threonine and tyrosine) are found scattered within other groups.

#### *1.2.2.2 TKL (Tyrosine kinase-like)*

Although, the sequences of TKL family members are similar to those of the tyrosine kinase family, these enzymes form a distinct, closely related grouping.

#### *1.2.2.3 AGC*

This group contains PKA, PKG, PKC as members. Therefore, this group is called the AGC family.

#### *1.2.2.4 CAMK (Calcium/calmodulin-dependent protein kinase)*

Calcium/calmodulin regulated kinases and structurally related kinase families

#### *1.2.2.5 CKI (Casein kinase 1)*

#### *1.2.2.6 CMGC*

This group contains CDK, MAPK, GSK3, CLK as members of the families. Therefore, it has this name from the initial letter of these kinases.

#### *1.2.2.7 STE (Homologs of yeast Sterile 7, 11, 20 kinases)*

This group comprises kinases in the MAP kinase cascade and homologs of yeast Ste7 (MAP2K), Ste11 (MAP3K) and Ste20 (MAP4K) kinases.

#### *1.2.2.8 RGC (Receptor Guanylate Cyclases)*

This group contains eukaryotic protein kinase domains that all appear to be catalytically inactive.

#### *1.2.2.9 AK (Atypical kinases)*

Atypical kinases consist of a relatively large family of important proteins such as PI3K, actin-fragmin kinase and choline kinase. They do not have significant sequence similarity to other eukaryotic kinase. Their structural character was studied in detail by Scheeff & Bourne (Scheeff & Bourne, 2005).

#### *1.2.2.10 Other*

Kinases that do not conform to any classification above.

Since Manning et al. published their phylogenetic tree, several human kinome-wide studies have mapped structural coverage (Fedorov et al, 2007; Marsden & Knapp, 2008) and binding interaction patterns (Fabian et al, 2005; Goldstein et al, 2008; Karaman et al, 2008) onto this human kinome tree. At the time of starting my PhD, the largest single source of publicly available kinase experimental binding data was the Supplementary Table 4 from Fabian *et al.* which had 119 kinases assayed against 20 kinase inhibitors using T7-phage expression assays (Fabian et al, 2005).

### *1.2.3 Architecture of the kinase fold*

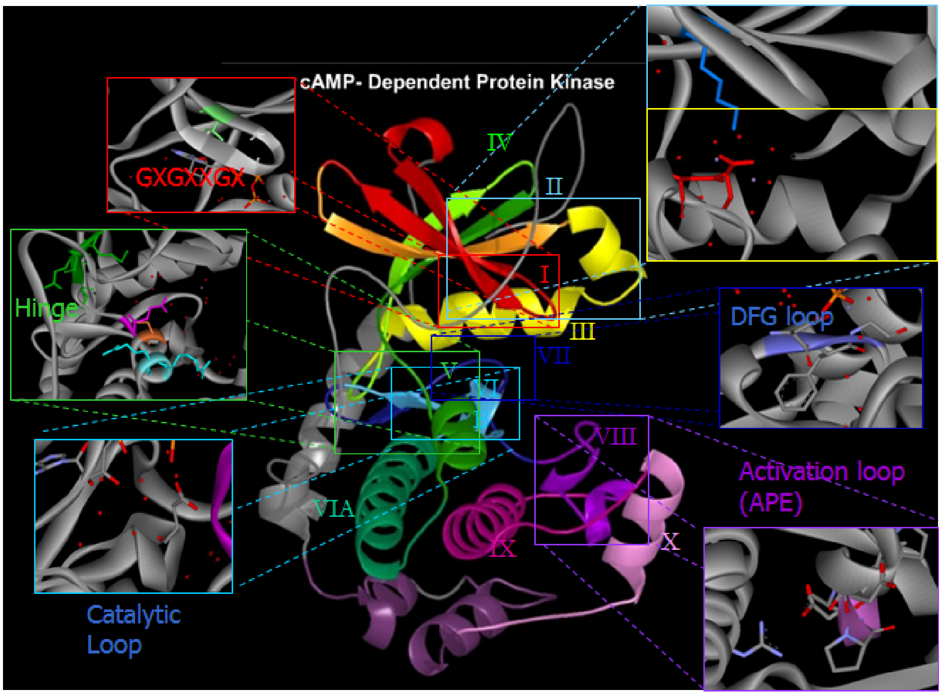


Figure 3. Subdomains of kinases exemplified by the structure of PKA (Niedner et al, 2006).

The kinase fold consists of the N-terminal lobe which contains mostly β-sheets and the C-terminal lobe which contains mostly α-helices. For more detail analysis, the kinase fold can be divided further into subdomains.

The descriptions of characters of each subdomain are taken from the structure walk-through of the protein kinase resource (Niedner et al, 2006) which follows the initial classification determined by Hanks & Hunter (Hanks & Hunter, 1995).

Table 1. Subdomain boundaries in PKA and CDK2

| General Info | PKA | CDK2 |
|---|---|---|
| Subdomain I | 43-64 | 4-25 |
| Subdomain II | 65-83 | 26-43 |
| Subdomain III | 84-98 | 44-58 |
| Subdomain IV | 99-113 | 59-73 |
| Subdomain V | 114-137 | 74-98 |
| Subdomain VIA | 138-160 | 99-121 |
| Subdomain VIB | 161-177 | 122-138 |
| Subdomain VII | 178-193 | 139-157 |
| Subdomain VIII | 194-210 | 158-175 |
| Subdomain IX | 211-240 | 176-209 |
| Subdomain X | 241-260 | 210-254 |
| Subdomain XI | 261-297 | 255-282 |

#### *1.2.3.1 Subdomain One:*

This subdomain contains the GXGXXGX motif, known as the glycine rich loop, which helps to anchor the ATP.

#### *1.2.3.2 Subdomain Two:*

There is an invariant lysine (Lys 72), which is shown to be very important for the catalytic activity of the enzyme

#### *1.2.3.3 Subdomain Three:*

This subdomain contains a prominent alpha helix (helix αC) (Huse & Kuriyan, 2002) and Asp 91 which form a salt-bridge with Lys 72.

#### *1.2.3.4 Subdomain Four:*

It is the beta strand IV, no invariant residues in this subdomain.

#### *1.2.3.5 Subdomain Five:*

This subdomain acts as a bridge to connect between 2 lobes. The hinge region (residue 81-84 in CDK2) contains a set of hydrogen bond donor and acceptor sites that is required for potent inhibitor binding (Davies et al, 2002). The hinge region begins with the gatekeeper residues, which form a hydrophobic pocket that surrounds the adenine ring. It contains Glu 127 which binds to Arg of pseudosubstrate.

#### *1.2.3.6 Subdomain Six A:*

This large helix plays a mainly structural role. There is no invariant residue.

#### *1.2.3.7 Subdomain Six B:*

This subdomain comprises of two small strands with a loop between them. This loop is called the catalytic loop because Asp 166 in the loop is likely to form the catalytic base that accepts the proton from the protein substrate's hydroxyl group.

#### *1.2.3.8 Subdomain Seven:*

This subdomain contains a conserved DFG motif. Asp 184 in the DFG motif helps in orienting the gamma phosphate of ATP by chelating the $Mg^{2+}$ ions.

#### *1.2.3.9 Subdomain Eight:*

This subdomain consists of the activation loop, which is used for substrate recognition and stabilisation. It contains a highly conserved APE motif. Many protein kinases become activated by phosphorylation of residues in this subdomain.

### *1.2.3.10 Subdomain Nine:*

This subdomain stabilises the catalytic loop and plays a role in pseudo-substrate recognition.

### *1.2.3.11 Subdomain Ten:*

This is a small alpha helix with unknown function.

### *1.2.3.12 Subdomain Eleven:*

This marks the C-terminal boundary of the kinase domain from the rest of the protein.

## *1.2.4 The ATP binding site*

According to the pharmacophore model for tyrosine kinases (Traxler & Furet, 1999), the ATP binding site can be divided into 5 regions which have distinct chemical environments (Figure 4).

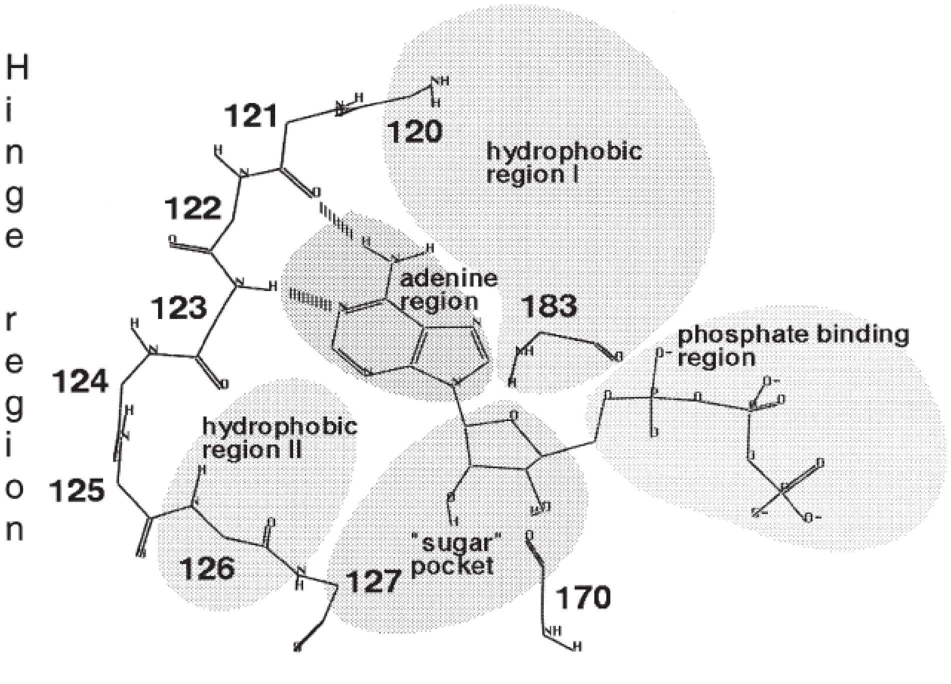


Figure 4. Protein kinase pharmacophore (Traxler & Furet, 1999). This residue numbering is for cyclic AMP dependent protein kinase (PKA). The GXGXXGX motif is the loop above the oxygen of ribose.

#### *1.2.4.1 Adenine region*

This is the region where all ATP-competitive kinase inhibitors bind. The major interaction of adenine in this hydrophobic pocket is the hydrogen bond donor acceptor system through the backbone carbonyl of Glu 121 and backbone NH of Val 123 in PKA. Although the backbone carbonyl of residue 123 is not used for ATP binding, it can serve as a hydrogen bond acceptor for some inhibitors, such as olomoucine (Schulze-Gahmen et al, 1995).

#### *1.2.4.2 Sugar pocket*

This region is often exploited to accommodate solubilising groups because of its hydrophilic character. It is not highly conserved and can be used to direct selectivity (Keri et al, 2006).

#### *1.2.4.3 Hydrophobic region I or hydrophilic backpocket*

This region is the space extending from the lone pair nitrogen of ATP. It is not conserved and has been used to gain affinity as well as selectivity for several potent inhibitors of serine/threonine and tyrosine kinase (Keri et al, 2006). Access to this pocket is controlled by an amino acid residue, which is equivalent to residue 120 in PKA, called the gatekeeper (Liu et al, 1999). There is evidence that the selectivity of pyrazolopyrimidines is controlled by the size of this amino acid (Schindler et al, 1999). Most tyrosine kinases have a small gatekeeper residue (threonine or valine), which makes them more sensitive to these drugs than most serine/threonine kinases, which have a larger gatekeeper residue (methionine or isoleucine). A mutant of tyrosine kinase Src with modified gatekeeper residue (T338I) has broad inhibitor resistance (Apsel et al, 2008).

#### *1.2.4.4 Hydrophobic region II or surface exposed front area*

This region is a hydrophobic slot open to solvent, which is not used by ATP (Traxler & Furet, 1999).

#### *1.2.4.5 Phosphate binding region*

This region has high solvent exposure and seems unimportant for binding affinity. It can be used to gain selectivity (Traxler & Furet, 1999).

Many more small pockets have been identified for the analysis of selective kinase inhibitors (Liao, 2007).

## 1.3 Binding site studies

There is evidence that high affinity targets sometimes have similar residues at positions important for binding of a given kinase inhibitor, although others with similar residues at these important positions can be insensitive to such inhibitors, probably due to conformational differences (Sheinerman et al, 2005). Therefore, understanding kinase selectivity cannot be achieved only through the analysis of sequences but must also consider three-dimensional structures.

Although biomolecules are in motion, we often treat them as static objects and determine their shapes by their surfaces. At the molecular level, the solid sphere representation is normally used to define shape (Morris et al, 2005). For relating biological activities between drug-like molecules and their molecular targets, the shape of the pocket is the most discriminating factor (Ballester & Richards, 2007).

Cavity detection methods have been divided into those that depend on energetic criteria and geometry criteria. An example of energetic-based detection is Qsitefinder (Laurie & Jackson, 2005) which calculates the van der Waals interaction energy of a methyl probe with the protein. The probes, which have favourable interaction energies, are obtained and a cluster of these probes is ranked according to interaction energy. This method has proven efficient in the detection of binding sites.

The geometry-based cavity detection methods have been divided further into volumetric based and surface based. Based on the description by Nayal and Honig, “Volumetric methods aim to identify spaces in the vicinity of the protein sequestered by protruding protein atoms, while surface-based approaches locate surface cavities by an analysis of the geometry of the molecular surface itself” (Nayal & Honig, 2006).

Early solvent accessible surface methods were developed by Lee and Richards (Lee & Richards, 1971) and later improved by Connolly (Connolly, 1983). The algorithm employed a sphere of solvent molecule to roll over the protein to generate a smooth surface. Surface-based approaches have been very useful for visualising shape-complementarity and protein-protein interfaces. Many site comparisons use volumetric-based methods, which can be divided further as to whether or not they rely on the 'grid method'.

If the protein is embedded in a 3D grid, points, which are not overlapped by protein atoms, are detected and then certain criteria are applied to judge whether they are parts of the pocket. Examples of programs that use the grid volumetric-based method are Ligsite (Hendlich et al, 1997) used in Cavbase (Kuhn et al, 2006) and VOIDOO (Kleywegt & Jones, 1994). If the protein is not embedded in a 3D grid, surfaces are either identified by Voronoi tessellations or sphere clusters such as SURFNET (Laskowski, 1995) used in calculating real spherical harmonic expansion coefficients which can be used as 3D shape descriptors (Karaman et al, 2008; Morris et al, 2005 ).

In this thesis, I describe surfaces defined by volumetric-based geometric criteria, either through frequently occurring atoms on a grid (Chapter 2) or at representative points without a grid (Chapter 3), in order to describe the overall ‘shape’ of the kinase pocket.

Due to their well-defined shapes and the pharmaceutical importance of protein kinases, ATP binding pockets have been used as case studies for intensive binding site analyses in several papers. In 2002, Naumann and Matter classified kinases based on similarity in protein-ligand interaction features in 26 X-ray structures from 6 kinase families. They used GRID force-field probes (i.e. hydrophobic probe, sp2 carbonyl oxygen probe, neutral flat NH probe) with 1 Å spacing to derive binding site information in the form of interaction energies on a GRID molecular interaction field. After that, the data matrix was scaled and principal component analysis was used to cluster kinases with common interaction patterns. The approach can classify kinase-binding sites into subfamilies and identify favourable interactions that are characteristic for each class.

In 2007, Kuhn et al. used Cavbase to classify protein kinases from 258 X-ray structures in 48 subfamilies (Kuhn et al, 2007). Cavbase uses Ligsite which generates grids with a 0.5 Å spacing to detect buried cavities. If there are atoms in any amino acid that are closer than 1.1 Å to a surface point, such atoms will be used as a pseudo centre to represent the chemical property of that region. Common substructures are identified by the clique algorithm (Bron & Kerbosch, 1973). The closest matches are compared by determining the overlap between surface patches and pseudo centres. Similarity matrices generated from various clustering methods were used as inputs for principal component analysis, enabling the detection of cross-reactivity between various kinases.

In 2009, Kinnings and Jackson used a geometric hashing algorithm to compare binding sites based on atom-atom similarity in an all-against-all manner amongst 354 crystal structures from 75 different kinases (Kinnings & Jackson, 2009). Their geometric hashing procedures can be divided into a pre-processing stage and a recognition stage. In the pre-processing stage, distance features between all atom pairs of the model molecules are calculated and converted into a hash table. In the recognition stage, the same representations are calculated for the target molecule and this information can be used to access the hash table. If the pattern of the second matches that of the first by more than a certain score, a rigid body transformation is applied to expand the number of matching points (Via et al, 2000). Similarity scores were clustered and comparisons were made between the sequence-based and binding site based classifications. The authors found that many clusters can be generated from related kinases, and hence structural similarity is sufficient for the classification of the highly conserved ATP binding site in kinases.

## 1.4 Quantitative Structure-Activity Relationship (QSAR)

This methodology has been developed following the idea of Crum-Brown and Fraser who proposed that the physiological action of a substance relates to its chemical constitution (Crum-Brown & Fraser, 1868) and independent reports which suggested that there is a linear relationship between the depressant action of organic compounds and their oil/water partition coefficient (Meyer, 1899; Overton, 1901). Later, Hammett defined a linear free energy relationship between the reactivity and electronic properties ($\sigma$) of aromatic substituents with their equilibrium constants (Hammett, 1935). This relationship formed the mechanistic basis for the development of the linear Hansch equation and then a more successful parabolic equation as shown below (Hansch, 1969; Hansch & Fujita, 1964).

Equation 1

$$\log (1/C) = a(\log P)^2 + b(\log P) + c\sigma + \ldots + k;$$

where $C$ is the molar concentration that produces a biological effect, $P$ is the octanol/water partition coefficient and $\sigma$ is the electronic Hammett constant. This improved equation which combines a parabolic model with various physicochemical properties allows for the description of structure-activity relationships that cannot be correlated with a single linear term.

A true structure-activity relationship model is the Free-Wilson approach (Kubinyi, 1993). Free and Wilson (Free & Wilson, 1964) described in the equation that biological activity can be defined as a contribution of the parent molecule (μ) and the group contribution of all structural features attached to that parent molecule ($a_i$) below.

Equation 2

$$\log (1/C) = \Sigma a_i + \mu$$

A combination of the Hansch and Free-Wilson equations leads to the mixed approach below:

Equation 3

$$\log (1/C) = a(\log P)^2 + b \log P + \mathrm{c}\sigma + \ldots + \Sigma a_i + k$$

This mixed approach equation is a more powerful tool for large and structurally diverse quantitative structure activity relationships (Kubinyi, 1993). More advanced QSAR methods make analogies of how a binding site feels the electrostatic potential of the ligand by mapping molecular fields calculated from atomic probes onto three-dimensional grids which include several thousands of points as descriptors (Kubinyi, 1998). This thesis explores the idea of relating biological activity with distance parameters using multiple linear regression in a similar manner to the QSAR method.

## 1.5 Fragment Recognition

After investigating cross-reactivity features, I explore the possibility of designing new staurosporine derivatives by utilising knowledge of fragment environments gained by using a web tool to assist in molecular design.

The fragment approach to molecular recognition is a very promising field in drug design (Fattori, 2004). A computational approach to modify the functional groups of compounds with topologically different scaffolds that exert the same biological activity, known as bioisosteric replacement or scaffold hopping, has also been the topic of several discussions (Zhao, 2007).

Several studies have searched for bioisosteres, functional groups that are structurally different but can form similar intermolecular interactions. A change of chemical template may give rise to a compound with better solubility, enhanced pharmacokinetic properties, improved binding affinity, and may lead to a novel compound which is patentable (Böhm et al, 2004). Some scaffolds have been shown to be active in several target proteins, so providing opportunities for use in different therapeutic areas. On the other hand, some scaffolds specifically target certain drug target families and can be used as molecular anchors for functional decoration (Muller, 2003).

The key methods in scaffold hopping include shape matching, pharmacophore searching, fragment replacement and similarity searching (Böhm et al, 2004). These methods differ in the way similarities between bioisosteric groups are determined (Wagener, 2006). Even though common chemical replacements in drug-like compounds have already been described (Sheridan, 2002) and a ring replacement database has been constructed (Lewell et al, 2003), the chemical environments of functional groups within protein structures have not yet been considered.

One interesting study exploited a set of crystal structures of proteins harbouring different ligands. Bioisosteres were identified from pairs of substructural features with high volume overlap from an overlay of the structures (Kennewell et al, 2006). Another study which used Relibase to find interacting functional groups and their preferred interaction geometries also pinpointed the importance of the chemical environment in the protein for identifying bioisosteres (Bergner et al, 2001). Such information about which set of atoms/residues interact with a certain fragment is not systematically incorporated in the available public databases.

Bemis analysed a set of 5120 drug molecules from the Comprehensive Medicinal Chemistry database and found that 50% of the drugs were derived from only 32 molecular frameworks (Bemis & Murcko, 1996). Similarly, Ertl analysed a set of 3 million compounds and identified the 50 most common substituents (Ertl, 2003). These sets of fragments give an estimation of the scale of the problem and can be a good starting point for a systematic analysis of the underlying structural basis of fragment recognition. If the Protein Data Bank contains enough molecular varieties, it might be possible to achieve bottom-up molecular design by linking together fragments that can fit into known chemical environment.

## 1.6 Thesis objectives

My aim was to develop computational approaches that could account for the specificity and discrimination of inhibitors for kinases and that might contribute to a general understanding of molecular recognition in other systems. Using the grid method, I developed a procedure to observe the region in the ATP-binding pocket which causes cross-reactivity among kinases. A method was developed that can compare the shapes of kinase pockets by using distance matrices measured from representative atoms. From these analyses, I rationalised how features with the most variable positions in the binding site affect binding affinity by using a multiple linear regression method. Lastly, I developed a tool to assist in the understanding of ligand binding and in minimising the choices for synthetic strategies to design specific kinase inhibitors.

# Chapter 2

## 2 CONSERVED ATOMS AND RESIDUES IN THE ATP BINDING SITE

*The conservation of atoms and residues in the active sites of kinases is analysed by superposing the structures of complexes with the same ligand and the entities that most often retain their position are observed. The neighbouring atoms surrounding the universal kinase inhibitor staurosporine have a higher degree of conservation in position than those surrounding adenosine complexes, which is consistent with the hypothesis that the rigidity of staurosporine can fix some parts of the active site in particular positions. The representation at residue level brings us closer to unraveling kinase specificity. It demonstrates that the C$\alpha$ of the first glycine of the glycine-rich loop (**G**XGXXG) is recruited towards the ether oxygen of the ribose upon staurosporine binding (found in 75% of the non-redundant staurosporine complexes). Therefore, the active site can contract or expand to accommodate a larger ligand, and staurosporine leads to an induced-fit in the pocket of the kinases.*

### 2.1 Introduction

The first phase of my research focused on an analysis of ATP binding sites in protein kinases because of their importance in chemotherapy and the availability of X-ray structures in the Protein Data Bank. I try to minimise the problems of conformational change, protein flexibility and electrostatic potential by focusing mainly on structures of protein kinases which bind to either adenosine derivatives or to a universal kinase inhibitor called staurosporine. This molecule is quite rigid, and only the parts that can make strongly electrostatic interactions mimic those of ATP. Hence, it should constrain the active site of those kinases that it binds sufficiently well for a comparison of surrounding atoms and binding affinities. I hypothesize that atoms that remain conserved both in atom type and position for complexes with the same ligand maintain their structural role to satisfy optimal interaction with the ligand. Therefore, surrounding kinase atoms which share a conserved position in both ATP and staurosporine complexes may be generally required for the binding of kinase ligands.

The idea of this approach to study position-specific interactions originated from image processing in electron microscopy where large numbers of images are superposed in order to intensify the true signal relative to the background noise. The algorithm developed constructs a grid box around the rigid part of superposed ligand and collects data points in four dimensions, i.e. residue or atom types, and x, y, z coordinates, from the non-redundant protein kinase structures. The final representation is obtained by converting the image to the form of a PDB file, in which signal-to-noise ratio thresholds can be altered in order to optimise visualization. The approach is encoded in new software called SIMFONEE (specificity implication from frequently occurring neighbouring entities).

## 2.2 Methods

### *2.2.1 Dataset*

The structures of protein kinases were selected from the MSDlite Database (Golovin et al, 2004) to allow queries from the Gene Ontology identification number (GO:0004672), which gives a precise identification for ‘protein kinase activity’. Structures with resolutions better than 3.0 Å and containing either staurosporine or adenosine phosphate moieties comprised the initial data set. Non-redundant PDB chains were selected manually to cover all staurosporine and adenine phosphate complexes. The structures of 20 staurosporine-kinase complexes were superposed on the indolocarbazole moiety from staurosporine in PDB ID 1stc (Prade et al, 1997) in order to compare them with the structures of 24 adenosine phosphate-kinase complexes, which were superposed on the adenine ring from ATP in PDB ID 1atp (Zheng et al, 1993). The chosen structures are listed in Table 2, along with details of their quality and the complex crystallised in Appendix A.

Table 2. PDB code and the chain used in position-specific study

| LIGAND | LIGAND CODE | PDB CODE FOLLOWED BY CHAIN IDENTIFIER |
|---|---|---|
| Staurosporine | STU | 1AQ10, 1BYGA, 1NVRA, 1NXKA, 1OKYA, 1Q3DA, 1QPDA, 1SM2A, 1STCE, 1U59A, 1XBCA, 1XJDA, 1YHSA, 2CLQA, 2DQ7X, 1WVYA, 2ITWA, 2OICA, 2GCDA, 2HW7A |
| Adenosine phosphate | ATP | 1ATPE, 1OL6A, 1QL6A, 1S9JA, 1QMZ 1U5RA, 1ZYDA, 2BIYA |
| | ANP | 1DAWA, 1IR3A, 1J1BA, 1JKLA, 1LP4A, 1MQBA, 1O6KA, 1PJKA, 1QPCA, 1YXTA, 2A19B, 1Q99 |
| | ACP | 1K3AA, 1O6YA, 1U54B |
| | AMP | 2IVTA |

### *2.2.2 Classification of neighbouring atom types*

The position-specific interactions were considered at two-levels: the atom type and the residue type. For atom matrices, atoms in the PDB file were assigned an atom type according to the simplified approach used in the AMBER force field, which has been developed specially for molecular mechanics calculations of proteins and nucleic acids (Cornell, 1995) (Supporting Information 1). The atom type categorisation is based on the assumption that atoms around the side chains that have the same functional group can be classified as the same atom type, e.g. carboxylate oxygens of Asp and Glu have sp$^2$-oxygen atom type**s.** By using this approach, we can capture similar interactions in the pocket made by the same part of the ligand. The available atom types in this classification program are described below.

- *Carbons*
  1. **CT**: any sp$^3$ carbon
  2. **C**: any carbonyl sp$^2$ carbon
  3. **CA**: any aromatic sp$^2$ carbon and $C_\varepsilon$ of Arg
  4. **CH**: all of histidine's aromatic carbons except for CD2 (*i.e.* sp$^2$ aromatic carbon in 5-membered ring with one substituent and next to nitrogen, or next to carbon and lone pair nitrogen, or next to two nitrogens)
  5. **CW**: tryptophan's carbon in connection with the 5-membered ring (*i.e.* sp$^2$ aromatic in 5-membered ring next to carbon and NH, or at junction of 5- and 6-membered rings, or next to two carbons, or sp$^2$ junction between 5- and 6-membered rings and bonded to CH and NH)
  6. **CX**: any other unidentified carbon including carbon from the bound ligand
- *Nitrogens*
  7. **N** : sp$^2$ nitrogen in amides
  8. **NC** : sp$^2$ nitrogen in aromatic rings (sp$^2$ nitrogen with hydrogen attached, sp$^2$ nitrogen in 5-membered or 6-membered ring with lone pair electrons)
  9. **N2** : sp$^2$ nitrogen of aromatic amines
  10. **N3** : sp$^3$ nitrogen
  11. **NX**: any other unidentified nitrogen including nitrogen from the bound ligand
- *Oxygens*
  12. **OH** : sp$^3$ oxygen in alcohols, tyrosine, and protonated carboxylic acids
  13. **O**: sp$^2$ oxygen in amides
  14. **O2** : sp$^2$ oxygen in anionic acids
  15. **OW**: oxygen in water

16. **OX**: any other unidentified oxygen

- *Sulphur*

17. **SH**: cysteine sulphur

18. **S**: methionine sulphur

19. **SX**: any other unidentified sulphur

- *Phosphorous*

20. **P**: phosphorus in phosphates

- *Halogens*

21. **X**: F, Cl, Br, I

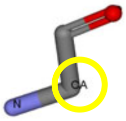
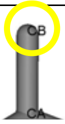
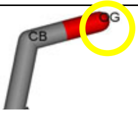
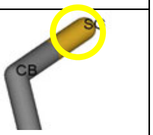
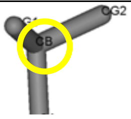
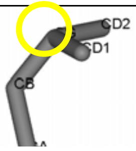
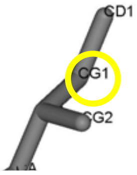
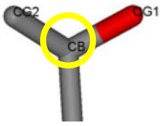
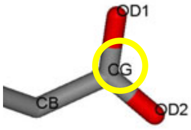
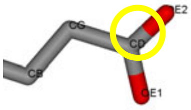
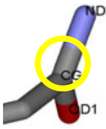
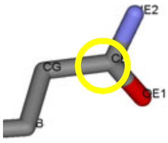
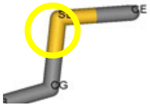
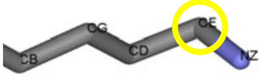
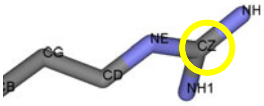
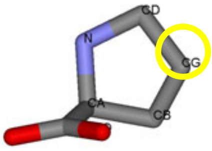
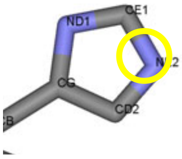
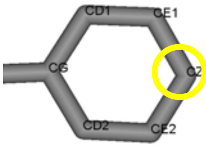
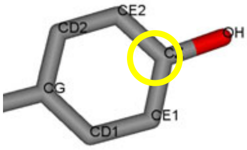
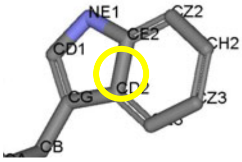
Table 3. Classification of amino acid atoms

| Atom | | Type |
|---|---|---|
| Glycine | N | N |
| Glycine | CA | CT |
| Glycine | C | C |
| Glycine | O | O |
| Alanine | N | N |
| Alanine | CA | CT |
| Alanine | C | C |
| Alanine | O | O |
| Alanine | CB | CT |
| Valine | N | N |
| Valine | CA | CT |
| Valine | C | C |
| Valine | O | O |
| Valine | CB | CT |
| Valine | CG | CT |
| Leucine | N | N |
| Leucine | CA | CT |
| Leucine | C | C |
| Leucine | O | O |
| Leucine | CB | CT |
| Leucine | CG | CT |
| Leucine | CD | CT |
| Isoleucine | N | N |
| Isoleucine | CA | CT |
| Isoleucine | C | C |
| Isoleucine | O | O |
| Isoleucine | CB | CT |
| Isoleucine | CG1 | CT |
| Isoleucine | CG2 | CT |
| Isoleucine | CD | CT |
| Serine | N | N |
| Serine | CA | CT |
| Serine | C | C |
| Serine | O | O |
| Serine | CB | CT |
| Serine | OG | OH |
| Threonine | N | N |
| Threonine | CA | CT |
| Threonine | C | C |
| Threonine | O | O |
| Threonine | CB | CT |
| Threonine | OG1 | OH |
| Threonine | CG2 | CT |
| Cysteine | N | N |
| Cysteine | CA | CT |
| Cysteine | C | C |
| Cysteine | O | O |
| Cysteine | CB | CT |
| Cysteine | SG | SH |
| Proline | N | N |
| Proline | CA | CT |
| Proline | C | C |
| Proline | O | O |
| Proline | CB | CT |
| Proline | CG | CT |
| Proline | CD | CT |
| Aspartate | N | N |
| Aspartate | CA | CT |
| Aspartate | C | C |
| Aspartate | O | O |
| Aspartate | CB | CT |
| Aspartate | CG | C |
| Aspartate | OD | O2 |
| Asparagine | N | N |
| Asparagine | CA | CT |
| Asparagine | C | C |
| Asparagine | O | O |
| Asparagine | CB | CT |
| Asparagine | CG | C |
| Asparagine | OD1 | O |
| Asparagine | ND2 | N |
| Glutamate | N | N |
| Glutamate | CA | CT |
| Glutamate | C | C |
| Glutamate | O | O |
| Glutamate | CB | CT |
| Glutamate | CG | CT |
| Glutamate | CD | C |
| Glutamate | OE | O2 |
| Glutamine | N | N |
| Glutamine | CA | CT |
| Glutamine | C | C |
| Glutamine | O | O |
| Glutamine | CB | CT |
| Glutamine | CG | CT |
| Glutamine | CD | C |
| Glutamine | OE1 | O |
| Glutamine | NE2 | N |
| Methionine | N | N |
| Methionine | CA | CT |
| Methionine | C | C |
| Methionine | O | O |
| Methionine | CB | CT |
| Methionine | CG | CT |
| Methionine | SD | S |
| Methionine | CE | CT |
| Lysine | N | N |
| Lysine | CA | CT |
| Lysine | C | C |
| Lysine | O | O |
| Lysine | CB | CT |
| Lysine | CG | CT |
| Lysine | CD | CT |
| Lysine | CE | CT |
| Lysine | NZ | N3 |
| Arginine | N | N |
| Arginine | CA | CT |
| Arginine | C | C |
| Arginine | O | O |
| Arginine | CB | CT |
| Arginine | CG | CT |
| Arginine | CD | CT |
| Histidine | N | N |
| Histidine | CA | CT |
| Histidine | C | C |
| Histidine | O | O |
| Histidine | CB | CT |
| Histidine | CG | CH |
| Histidine | ND1 | NC |
| Histidine (HE/+) | CD2 | CW |
| Histidine (HD) | CD2 | CH |
| Histidine | CE1 | CH |
| Histidine | NE2 | NC |
| Phenylalanine | N | N |
| Phenylalanine | CA | CT |
| Phenylalanine | C | C |
| Phenylalanine | O | O |
| Phenylalanine | CB | CT |
| Phenylalanine | CG | CA |
| Phenylalanine | CD | CA |
| Phenylalanine | CE | CA |
| Phenylalanine | CZ | CA |
| Tyrosine | N | N |
| Tyrosine | CA | CT |
| Tyrosine | C | C |
| Tyrosine | O | O |
| Tyrosine | CB | CT |
| Tyrosine | CG | CA |
| Tyrosine | CD | CA |
| Tyrosine | CE | CA |
| Tyrosine | CZ | C |
| Tyrosine | OH | OH |
| Tryptophan | N | N |
| Tryptophan | CA | CT |
| Tryptophan | C | C |
| Tryptophan | O | O |
| Tryptophan | CB | CT |
| Tryptophan | CG | CW |
| Tryptophan | CD1 | CW |
| Tryptophan | CD2 | CW |
| Tryptophan | NE1 | NC |
| Tryptophan | CE2 | CW |
| Tryptophan | CE3 | CA |
| Tryptophan | CZ2 | CA |
| Tryptophan | CZ3 | CA |
| Tryptophan | CH2 | CA |
| Arginine | NE | N2 |
| Arginine | CZ | CA |
| Arginine | NH | N2 |

### *2.2.3 Classification of neighbouring residues types*

Residue classification in SIMFONEE includes 20 amino acids and water, which comprises the 21 residue types. Cystine is omitted because it occurs rarely inside the cell. Each amino acid residue is simplified to one point at the position of its representative atom located at a distinctive part near the end of the side chain. The coordinate of this centre of the residue is a representative position of the whole amino acid at the residue level.

Table 4. Representative atoms of the 20 amino acids

| **GLY CA** | **ALA CB** | **SER OG** | **CYS SG** |
|---|---|---|---|
|  |  |  |  |
| **VAL CB** | **LEU CG** | **ILE CG1** | **THR CB** |
|  |  |  |  |
| **ASP CG** | **GLU CD** | **ASN CG** | **GLN CD** |
|  |  |  |  |
| **MET SD** | **LYS CE** | **ARG CZ** | **PRO CG** |
|  |  |  |  |
| **HIS NE2,** | **PHE CZ** | **TYR CZ** | **TRP CD2** |
|  |  |  |  |

### *2.2.4 Algorithm of the software SIMFONEE*

This software written in Perl derives structure-based information from crystal structures of protein kinases. I have developed a computer program to extract generalized features that are frequently found in protein kinase structures by constructing a 4-dimensional grid to capture different entities that are conserved in atomic position on structure superposition of staurosporine complexes. The grid collects occupancies from atoms from superposed structures that satisfy the four criteria, i.e. x,y,z coordinates and atom or residue type.

#### *2.2.4.1 PDB file preparation*

The input to this program is a list of PDB chain codes. The program downloads the PDB file from a local repository and extracts only the structure from the specified chain, or takes the whole structure if it is defined as having chain code 0. This structure is processed by program Moleman2 in order to set all the water molecules to chain ‘W’, and all the ligands to chain ‘Q’. Proteins without chain letters are reassigned by the program PDBset to have chain letter Z. The objective of this process is to avoid ambiguities arising from having several atoms with the same sequence number and the same chain which may confuse the classification and superposition programs. The processed structures are written to a new filename with a 4 letter PDB code and one extra-letter to identify the selected chain.

#### *2.2.4.2 Superposition on the ligand*

In order to allow visualization in the same coordination system, all output files should be in the same orientation. Therefore, all other PDB chains are superposed onto the same set of templates and kept in separate directories according to the template used for superposition. The catalytic subunit of bovine cyclic 3'- 5' adenosine monophosphate-dependent protein kinase (PKA) in complex with staurosporine, PDB ID 1stc (Prade et al, 1997), is used as the template and the reference residue nomenclature in this study because PKA is the first protein kinase for which a crystal structure became available and its residue nomenclature is widely used in kinase analyses. Therefore, the several templates for superposition are oriented according to this structure in order to refer to its residue numbers. The staurosporine template was taken from residue STO of the staurosporine in PKA (1stcE). Similarly, the template for the adenosine ring was taken from residue ATP of PKA structure 1atp.pdb after the active site had been superposed onto 1stc.pdb so that the neighbouring environment observed from both templates are in the same frame of reference as shown in Figure 5.

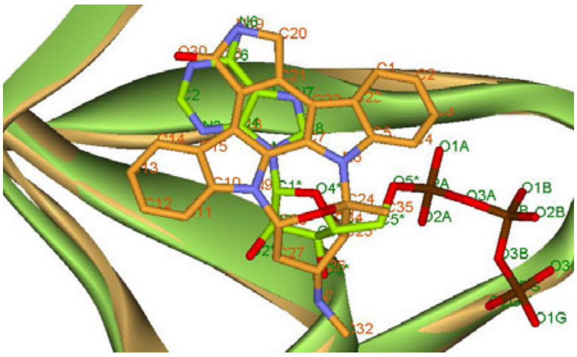


Figure 5. Staurosporine (carbons in orange stick) in complex with PKA, with the main chain in the same orientation as that found in the complex of PKA with ATP (carbons in green sticks)

For every input PDB file, SIMFONEE scans for the ligand centre and extracts all neighbouring atoms within a 20 Å radius from the centre of the ligand. Then it superposes the ligand sphere onto the appropriate ligand template using the CCP4 program LSQKAB.

### *2.2.4.3 Superposition on the kinase subdomain*

Because the N-terminal lobes of protein kinases are quite plastic, the frequencies of residues from this domain obtained from main chain superposition would fluctuate in accordance with the overall conformation of the protein. Partial secondary structures of the protein PDB ID 1b39 (chain A) are used as the template for superposition to observe conserved residues when there are large conformational changes between the domains. The template for the N-terminal lobe superposition includes a major part of the β-sheets down to the hinge region and helix αc (Figure 6), whereas the catalytic loop and four other helices are used as the template for C-terminal lobe superposition (Figure 7).

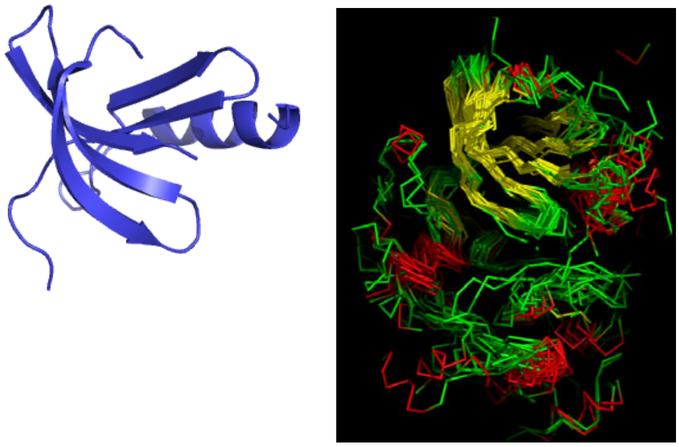


Figure 6. The N-terminal template (left) and superposition of the N-terminal domain (right)

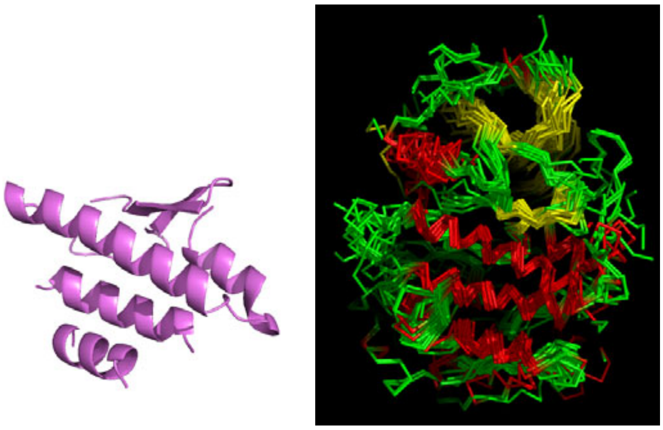
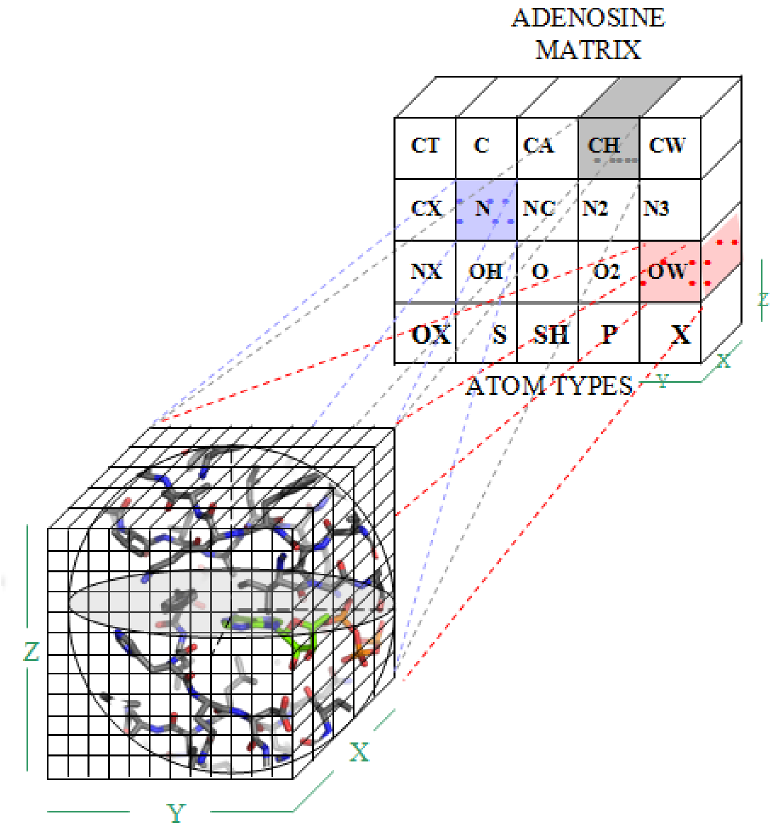


Figure 7. The C-terminal template (left) and superposition of the C-terminal domain

### *2.2.4.4 Counting frequently occurring atoms and residues*

The arrays of 4 dimensions were created for collecting the entities neighbouring the adenosine ring and staurosporine. Each block of the ligand array is defined by [type][x][y][z] and acts as a large bin to store occupancies from every surrounding atom which have the four parameters defined, i.e. type and x, y, and z coordinates. The entity is classified based on the either atom type or residue type as described earlier for the classification of the entities. The last three parameters describe the estimated position of the points at 1 Å step size by using the rounded integer of the coordinates x, y and z.



| | | | | |
|---|---|---|---|---|
| CT | C | CA | CH | CW |
| CX | N | NC | N2 | N3 |
| NX | OH | O | O2 | OW |
| OX | S | SH | P | X |

Figure 8. Illustration of filling the occupancies of the matrix identified by 4 parameters: [types][x][y][z].

### *2.2.4.5 Writing output by modifying the PDB format file and visualisation in Pymol*

For the atomic level, the grid was stored in PDB file format and occupancies of boxes were contoured using the color_b module of the program Pymol (DeLano, 2002) to obtain a transparent surface with the intensity of the colour corresponding to the frequency with which the grid boxes are populated. To visualise only the top rank entities, the grid can be displayed at different cut-off values. The array of frequently occurring atoms can also be displayed as a non-bonded sphere (Figure 9) with a label corresponding to the frequency with which the array is populated.

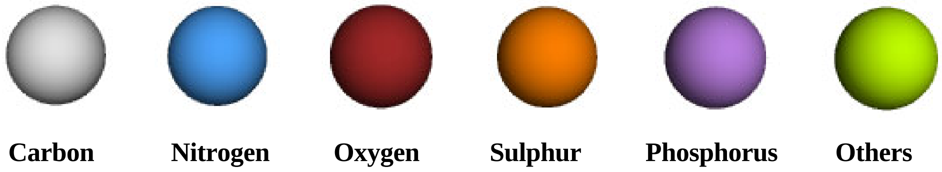


Figure 9. Non-bonded sphere representation for neighbouring atoms in the array

For the residue level, relative positions of neighbouring residues for different ligands were observed by superposing the majority of the residue clusters surrounding adenine and staurosporine, and then comparing the positions of these clusters for adenine complexes and staurosporine complexes. The staurosporine complexes are displayed with large sticks and adenosine phosphate complexes are displayed with small sticks, so that the relative positions of residue clusters can be compared. For visualising the cluster, a bond is drawn automatically when two atoms come closer than about 2 Å so the cluster can be easily found. Colours of frequently occurring residues are assigned according to the type of the amino acid found in the array. (Figures 10 & 11).

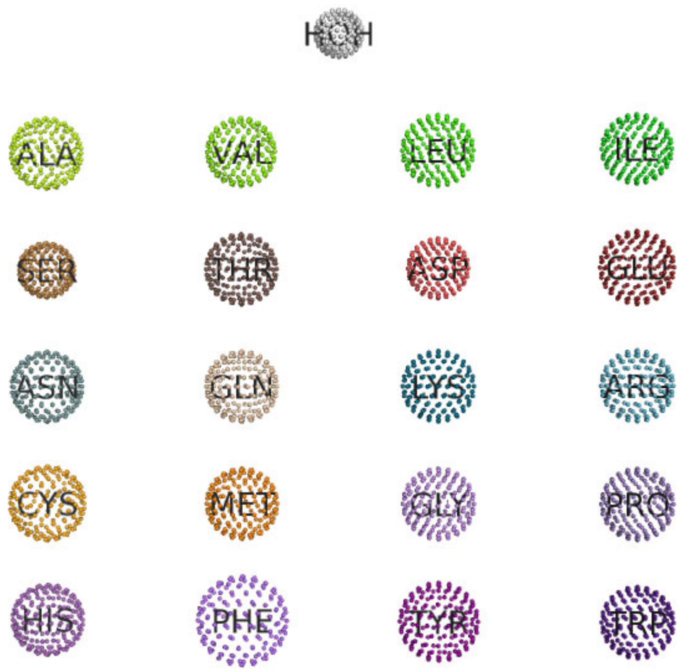


Figure 10. Water and amino acid neighbouring residues in dot representation

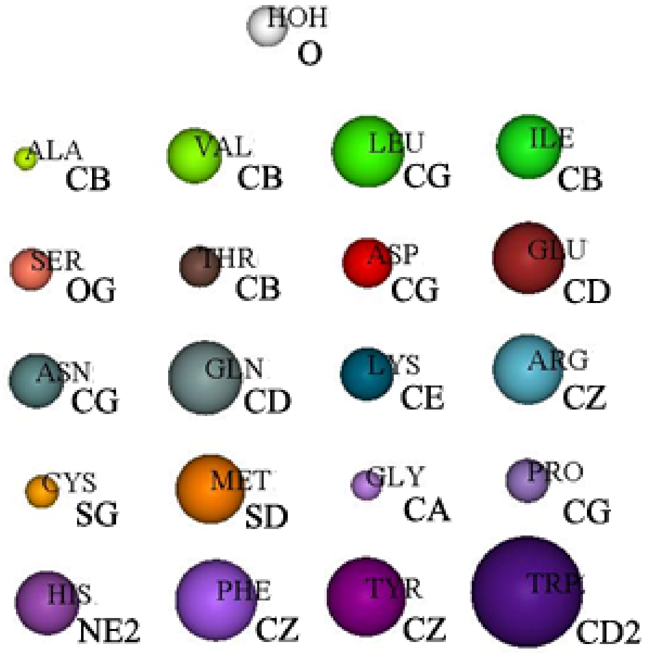


Figure 11. Water and amino acid residue in sphere representation. These atoms are used to represent the centres of the residues.

## 2.3 Results & Discussion

The staurosporine molecule is quite rigid as it contains very few rotatable bonds; hence we may observe interaction partners that are position-specific by superposing the kinases onto its lactam and indolocarbazole rings (20 structures of staurosporine complexes). In a similar manner interactions around the adenosine phosphate complexes can be compared by superposing the non-redundant kinase structures onto the adenine ring (24 structures of adenine-containing complexes). The conserved atomic environment can be found by observing the frequently occurring atoms at a particular location defined by a 1 Å grid box

### *2.3.1 The frequently occurring entities obtained by superposing ligand*

I observe that some clusters of amino acids preserve their functional groups in most of the staurosporine and the adenine complexes, for instance the side chains of Glu and Asp of the salt-bridges which flank each side of the pocket, the residues that are equivalent to Ala 70 in PKA which acts as the ceiling of the cleft, and Gly 50 which interacts with the ether oxygen of the sugar moiety (Figures 12 & 13). For the adenine complex, the most conserved part is the protein kinase main chain in the hinge region which interacts with the amine group of the adenine ring (Figure 12).

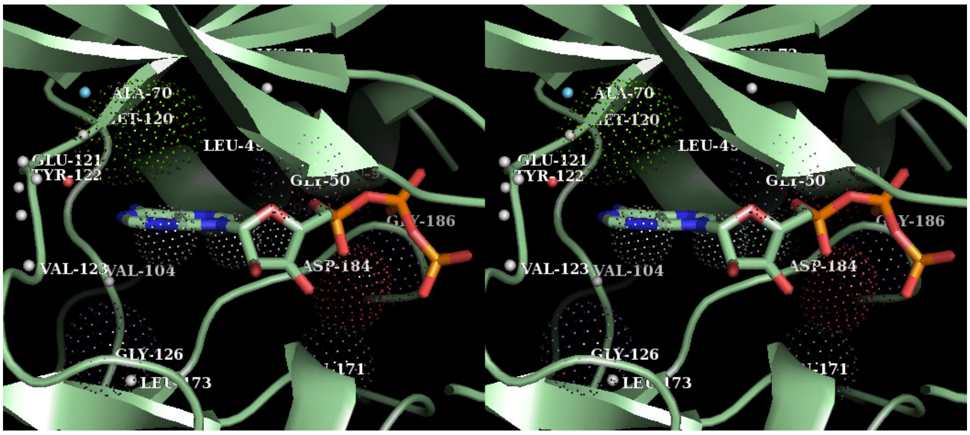


Figure 12. Stereo image illustrating frequently occurring neighbouring environment of adenosine ring at atomic level (spheres) and residue level (dots). The colour of atoms shown in Figure 9 and the colour of residues are shown in Figure 10.

For the staurosporine complex, the most conserved parts are the main chain of the hinge region, which interacts with the lactam oxygen, and the α-carbon of the first glycine of the GXGXXG motif, which interacts with the sugar moiety (Figure 13).

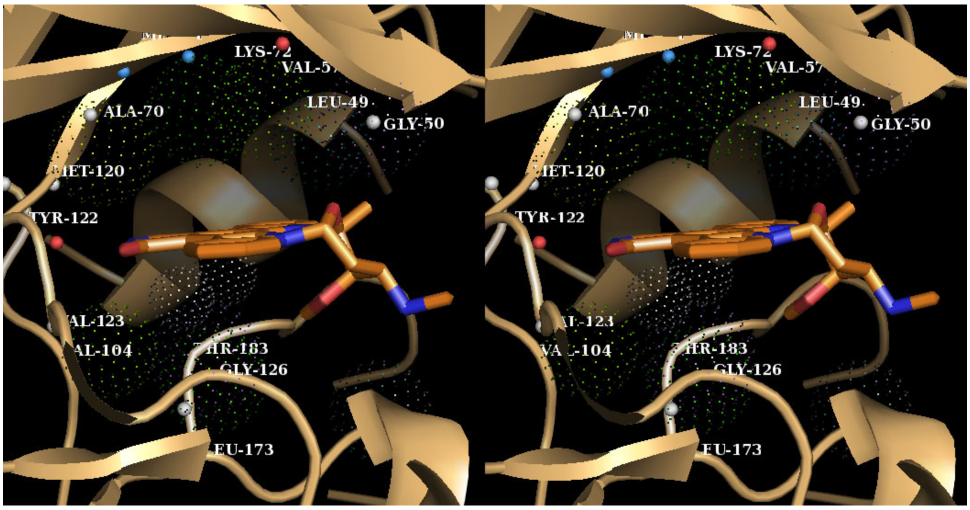


Figure 13. Stereo image illustrating frequently occurring neighbouring environment of staurosporine at atomic level (spheres) and residue level (dots). The colours of atoms are shown in Figure 9 and those of residues in Figure 10.

### *2.3.2 Comparison of flexibility of the pocket*

The resulting atomic level matrices suggest that for both adenine and staurosporine ligand complexes (Figures 14 & 15), the main chain atoms make the most conserved interactions in terms of type and position, which explains why staurosporine, mimicking ATP, can bind to most of the kinases.

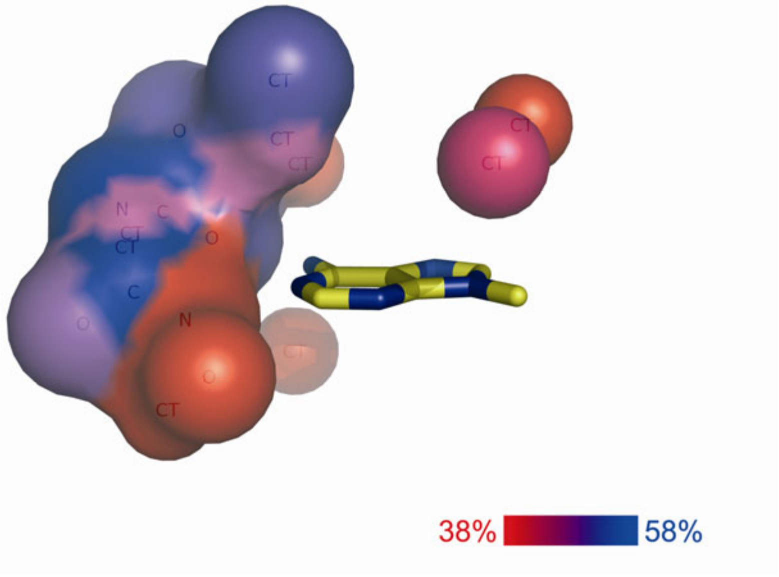


Figure 14. The hinge region (left) of the adenine phosphate binding complexes is the most conserved region. The colour of the surface is related to the occupancy of atoms in the array. The atom types (CT, N, C, O) are defined in Section 2.2.2.

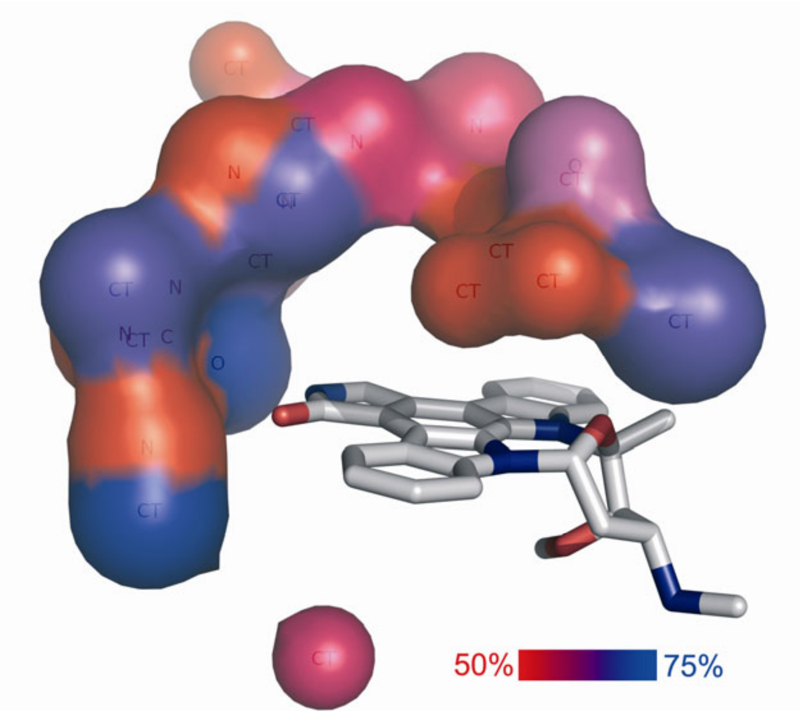


Figure 15. The maximum percent of conservation of frequently occurring neighbouring atoms in staurosporine complex (75%) shows that staurosporine pocket is more rigid than that of ATP. The atom types (CT, N, C, O) are defined in Section 2.2.2.

The conserved neighbouring atoms of ATP are more variable in position than those of staurosporine (37.5-58% conservation in Figure 14 versus 50-75% in Figure 15). This supports the idea that the two ligands require a different degree of flexibility within the active site of the kinase. The greater number of rotatable bonds in the adenosine phosphate results in lower degree of conservation of neighbouring atoms in these complexes than in staurosporine complexes. Since the ribose moiety of the adenosine complex can adopt several conformations, the frequently occurring atoms fall in several grid boxes.

### *2.3.3 Amino acid movement and substitution in the pocket*

The residue arrays serve to complement the pictorial representations of the atomic environment. For most of the residues, I chose the penultimate atoms of the side chain to represent the identities and the positions of the residues (Table 4). In this way, residues with a similar functional group at the end of the side chain in different kinases can be captured as points at a similar location in the superposed structures. For instance, $C_\beta$ of valine and $C_\gamma$ of leucine at the active sites occupy the same or close-by grid boxes in the superposed structures. Several clusters of amino acids are seen to have moved, leading to contraction and expansion of the residues in the pocket to accommodate staurosporine (Figure 16).

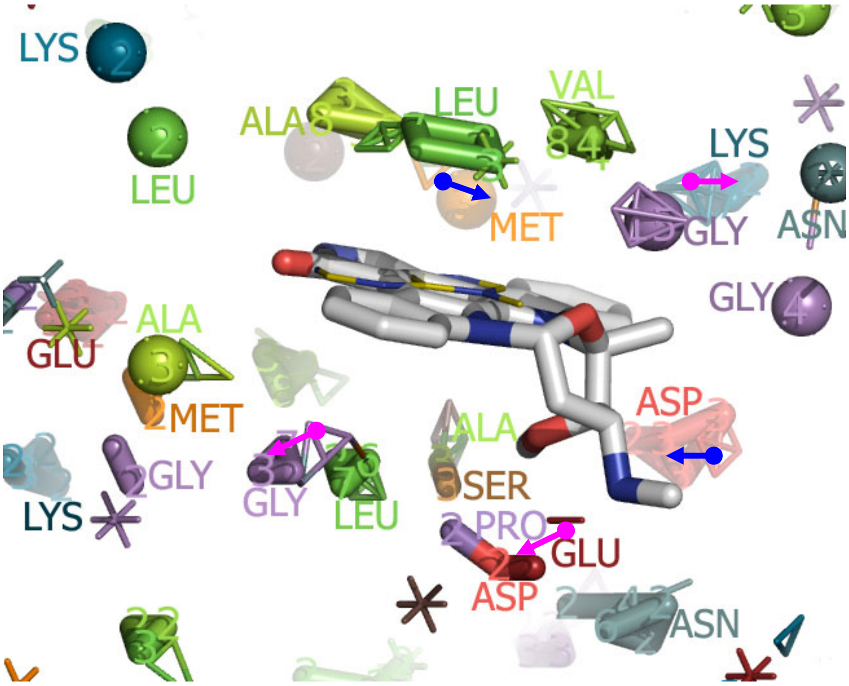


Figure 16. Contraction (blue arrows) and expansion (pink arrows) of the staurosporine pocket found when the residues in the environment of staurosporine (white large sticks) are superposed on those of adenine phosphates (yellow thin sticks).

The ends of some similar hydrophobic side chains e.g. Cys and Met, or Ala, Val and Leu, which surround the planar indolocarbazole ring, have equivalent positions implying that these amino acids perform the same function in that part of the active site. On the other hand, the amino acids that make contact around the methyl amino and methoxy groups of staurosporine demonstrate that remarkably different functional groups can occupy the same position and carry out the same structural role.

### *2.3.4 Induced fit caused by staurosporine*

Seventy five percent (15 out of 20 staurosporine bound structures) of the main chain alpha carbons from the first glycine of the glycine rich loop, Gly 50, fall in the same 1 Å grid-box (Figure 17). On the other hand, these atoms are distributed within a rhombohedron-shaped volume (see Figure 18) when superposed on the adenine ring.

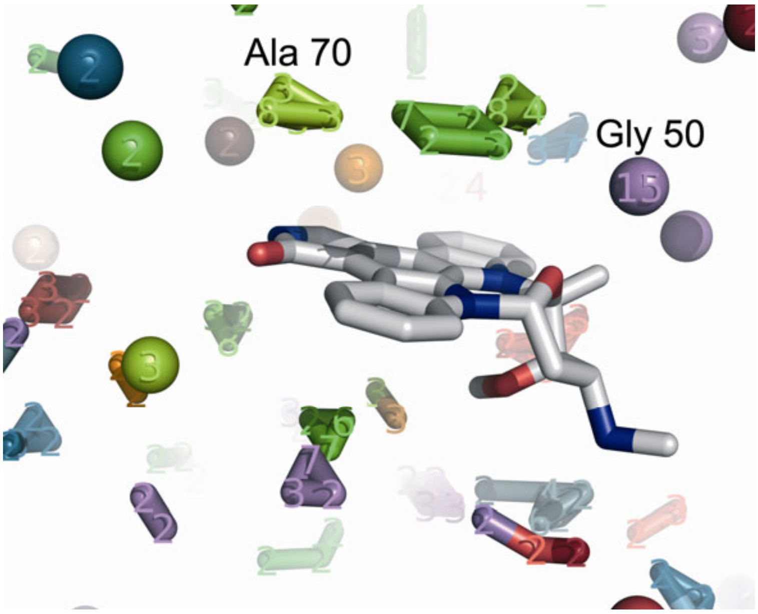


Figure 17. The first glycine of the GXGXXGX motif (Gly 50) is found in the same grid box in fifteen from twenty staurosporine binding structures. The residue that is most conserved in position during N-terminal domain superposition is Ala 70.

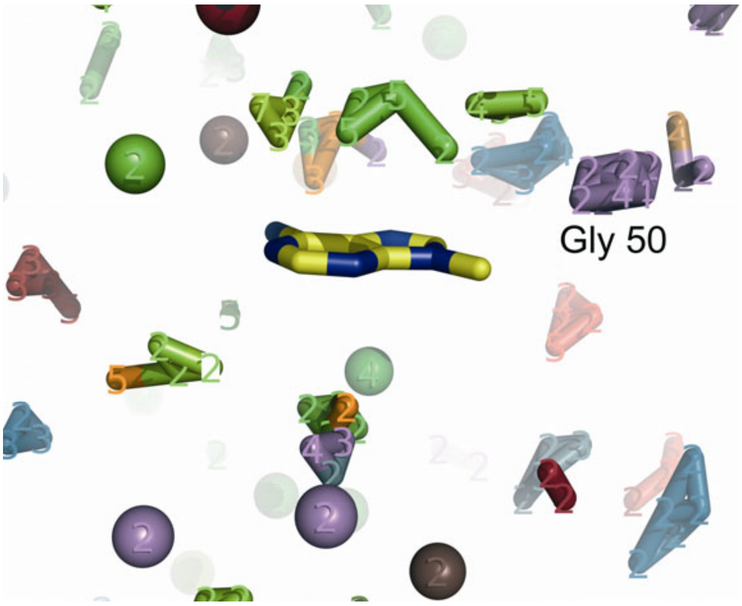


Figure 18. The first glycine (Gly 50) of the GXGXXGX motif can move freely in a rhombohedron-shaped volume when surrounding adenosine phosphates

This is perhaps surprising as the glycine-rich loop is generally believed to be highly flexible because of the displacement observed in crystal structures at pHs lower than 6 (Hemmer et al, 1997; Sachsenheimer & Schulz, 1977). However, this glycine becomes fixed in position upon staurosporine binding. The well-conserved position of this glycine, for 75% of the non-redundant staurosporine complexes, but not for adenine complexes, suggests that the staurosporine leads to an induced fit or conformational selection in the kinases upon binding.

### *2.3.5 The frequently occurring atoms obtained by superposing the domain*

When superposing neighbouring environments onto domain templates, I hypothesised that the residues that occur most frequently are the most rigid parts of the domain resulting from atoms that move in the same direction and with the same magnitude. The resulting amino acid residues from the N-terminal domain are not surprising. More than half of the 44 structures have their main chain sp$^3$ carbons (28 structures) and nitrogens (25 structures) nearby Ala 70 confined within a single element of the 1 Å grid (Figure 19). Both of these main chain atoms precede Ala 70 which acts as the ceiling of the pocket (Figure 20). Furthermore, the side chain of Val 57 which locked the adenine ring into place is presented as the most frequently occurring from the residue array (Figure 20).

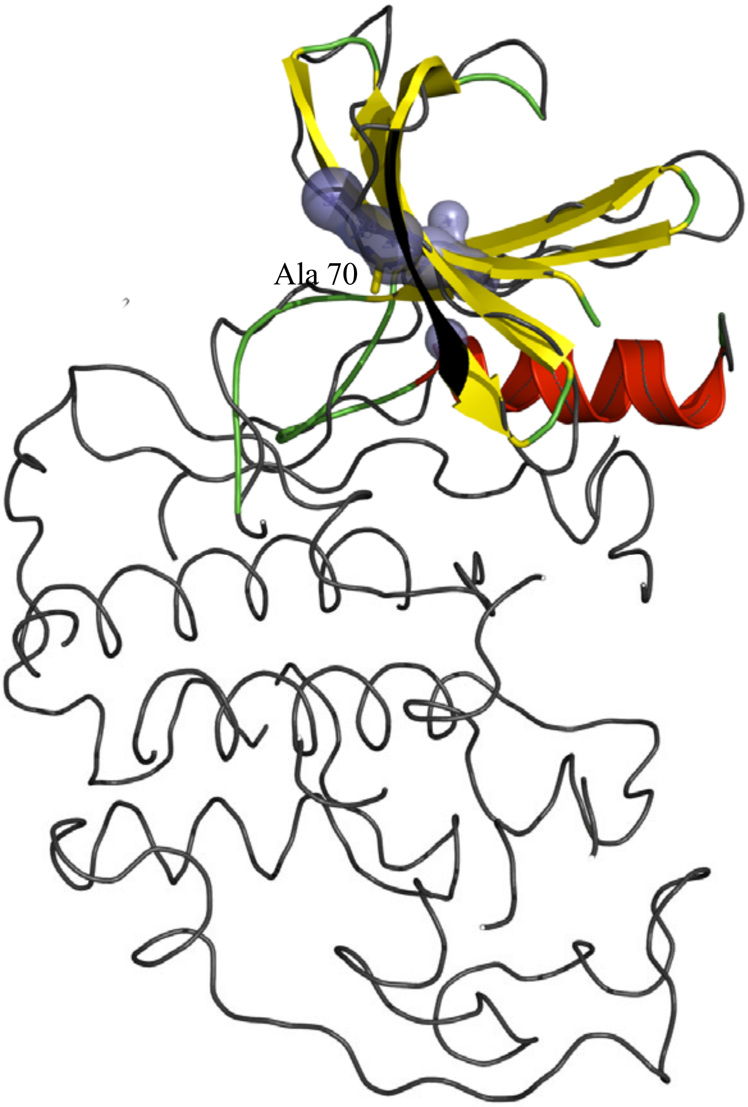


Figure 19. Blue surfaces are atoms that frequently retain their positions from the domain superposition using the N-terminal domain of CDK2 as a template. The template's helix is in red and sheets are in yellow. Ala70 is in yellow stick.

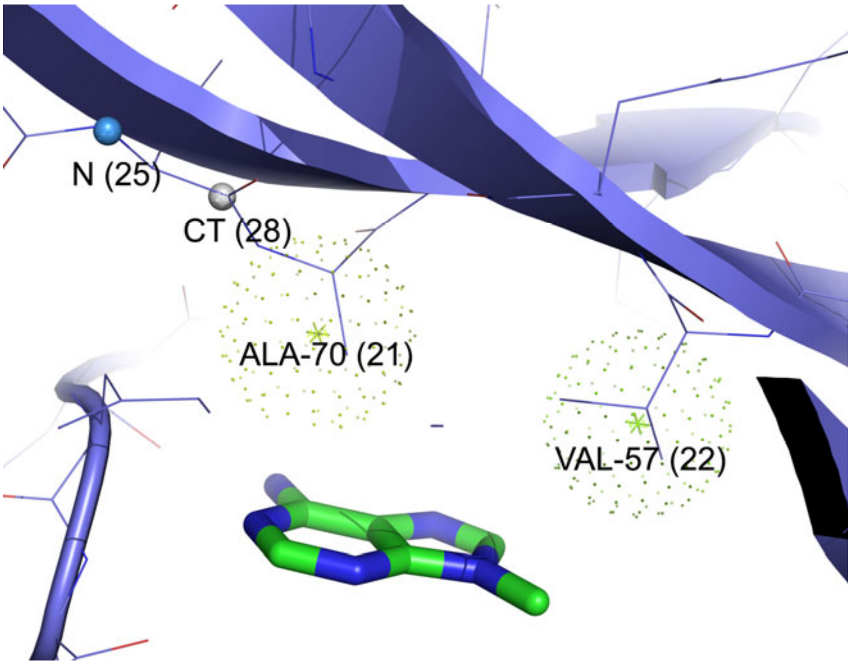


Figure 20. N-terminal domain superposition reveals the frequently occurring atoms and residues, with occupancies in brackets, from the total of 44 superposed structures.

The picture obtained from C-terminal domain superposition is more complicated. The side chain and the main chain of Asp 166 are among the most populated entities in the array (Figure 22). This residue is neither buried nor part of a helix or sheet. It appears that this aspartate interacts with a conserved His 125 in the CDK2 template, i.e. Tyr 164 in PKA, which is located just under the DFG loop (Figure 22). In addition, the beta carbon of Trp 221 often retains its position because it is next to the invariant Asp 220 that acts to stabilize the catalytic loop by hydrogen bonding to the backbone amides of, again, Tyr 164 at a few positions preceding the beginning of the catalytic loop. These frequently occurring atoms and residues support the idea that the catalytic loop is actually rigid. If it can move, it should move in the same way as the C-terminal domain that is used as the template for superposition. The most conserved residue is Gly 225, which locates in the middle of the helix, perhaps as a result of the superposition process.

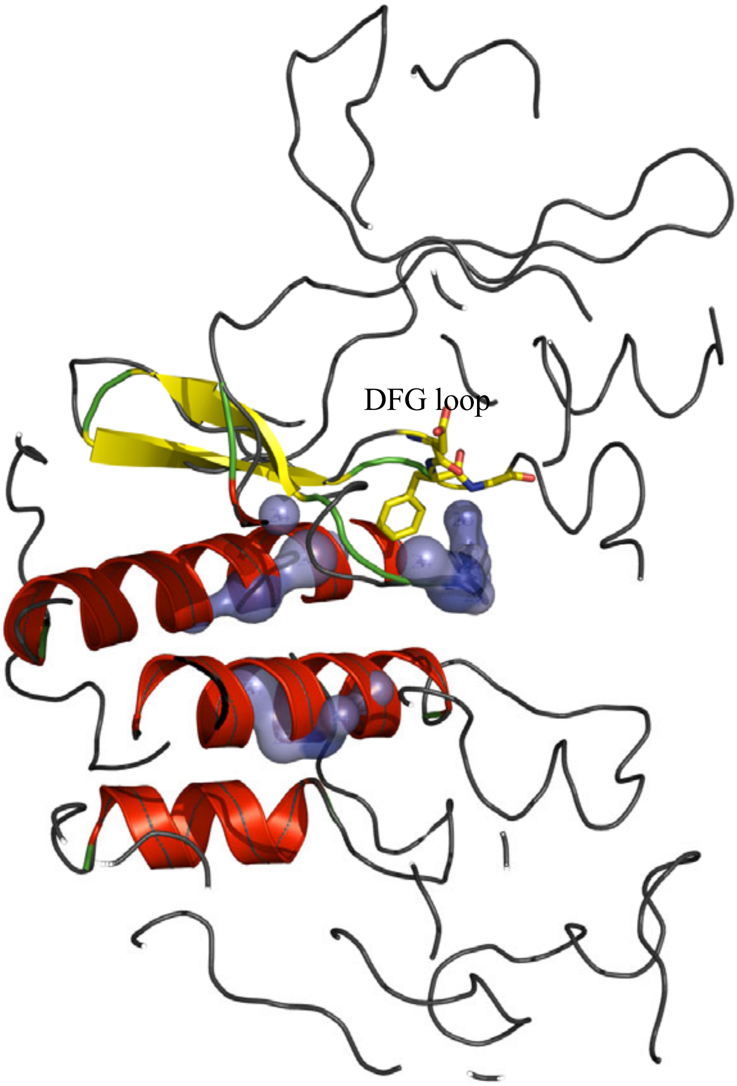


Figure 21. Blue surfaces are atoms that frequently retain their position from the domain superposition using the C-terminal of CDK2 as a template. The DFG loop is shown in yellow stick.

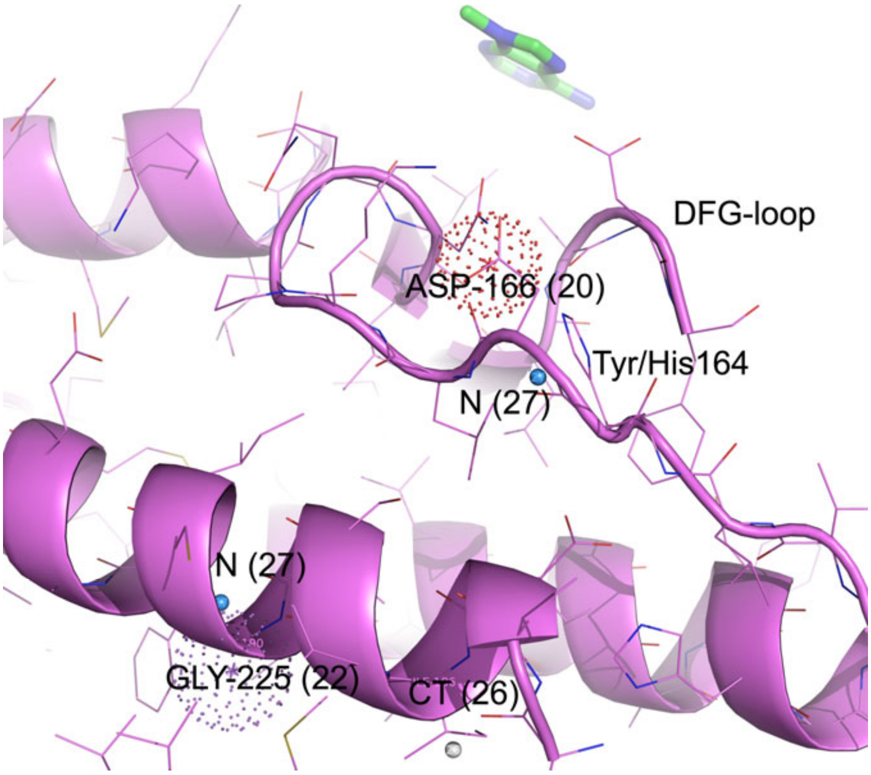


Figure 22. C-terminal domain superposition illustrates conservation of both main chain and side chain of PKA equivalent residue Asp 166 and Gly 225, and beta carbon of residue Trp 221 which is equivalent to CT (26) of Ile in this CDK2 structure. The frequently occurring atoms are in spheres and frequently occurring residues are in dots, with occupancies in brackets, from the total of 44 superposed structures.

## 2.4 Conclusion

The representation of conserved atoms portrays the rigidity of structural units within the ATP binding pocket as well as those distant from the pocket. Both the hinge region and the catalytic loop serve as crucial scaffolds conserving the positions of residues through the evolution of protein kinases. Representations of conserved residues demonstrate that there are adaptations of the ATP binding pocket upon ligand binding and that staurosporine causes an induced fit of the GXGXXGX motif. These subtle differences are observable when comparing the frequently occurring neighbouring residues of ATP with those of staurosporine. The generalisations do not conform to the general belief that the glycine rich loop and the catalytic loop are flexible. When the structures of kinases bound with either adenosine phosphate or staurosporine are superposed on a template of the N-terminal lobe, Ala 70 in the middle of the lobe retains its position. Alanine does not have strongly electrostatic side-chain features and it is interesting to know that it is held in a very precise position in the middle of this very labile secondary structure, the N-terminal lobe of the kinase. Furthermore, using all the large helices in the C-terminal lobe as the template showed that the catalytic loop was held in place with support from the helix in subdomain IX. From this approach, both the glycine rich loop and the catalytic loop have very precise locations. It is tempting to believe that in kinases with ligand bound, these two loops move with the secondary structure rather than swinging around.

However, this is a difficult concept to prove because the method relies very much on the part of the structures that are chosen as templates. It also depends on the superposition algorithm which atoms the program prefers to align with the template.

The next problem worth investigating is the peculiar spatial conservation of the penultimate atoms, which retain their positions in the C-terminal lobe superposition. The spatial arrangement of these atoms in the active site using atoms either at the penultimate or the end of the residues may give interesting clues about the basis of inhibitor selectivity.

# Chapter 3

## 3 SHAPE COMPARISON OF THE ATP BINDING SITE

*The positions of residue clusters are exploited to define the shape of the ATP binding site in protein kinases and to monitor the expansion and contraction of the pocket between different kinases and on ligand binding. Kinases with a similar spatial arrangement of residues have been shown to bind to a similar set of inhibitors. An analysis of the correlation of distance matrices demonstrates that there is no relationship between the similarity in amino acid sequence and the similarity in the spatial arrangement of the side chains in the pocket. On the other hand, kinases with high similarity in the spatial arrangement of the side chains tend to bind a similar set of inhibitors.*

### 3.1 Introduction

Current global protein structure classification schemes such as CATH and SCOP do not adequately underpin drug discovery because they do not provide information about binding site similarities (Debe & Hambly, 2004). In principle such information should give clues to potential targets and mechanisms of action, so assisting the design of selective therapeutic compounds (Jacoby, 2006). In this chapter, I explore the possibility of relating sequence and structural information to the ability to bind a ligand. Initially, I investigate whether kinases with similar binding site amino acid sequences tend to have the same binding affinities for a particular ligand. Secondly, I employ a novel shape comparison method to see whether kinases with similar shapes of their adenine binding site pockets have the same binding characteristics. Finally, I analyse dendrograms derived from both sequence and structural features to define relationships with real experimental binding data.

My shape comparison method relies on estimating correlations between distance matrices using the Mantel test, where the distances are measured from the representative, mostly penultimate, atoms of the residues. The challenge arises from the fact that the distances in the matrices are not independent of each other: changing the position of one object would change the distance from that object to each of the others. Therefore, the relationship between two matrices cannot be assessed by evaluating the correlation coefficient and testing its statistical significance. The classical Pearson correlation coefficient is used for measuring the correlation between the matrices, and the significance of this statistic is assessed by comparison with the distribution found by randomly reallocating the order of the elements in one of the matrices many times (Bonnet & Peer, 2002). Therefore, this procedure can overcome the problems arising from the statistical dependence of elements within each of the two matrices.

## 3.2 Method

### *3.2.1 Dataset*

The preliminary data sets used for residue selection were those structures from the PDB with IDs indicating ‘protein kinase activity’, i.e. those with gene ontology ID GO:0004672 in the MSDlite database (Golovin et al, 2004). These PDB IDs were filtered through the PISCES server (Wang & Dunbrack, 2003) to select a protein chain based on resolution, R-factor and completeness.

### *3.2.2 Kinase catalytic domain superposition and alignment*

The eighty chosen chains with resolutions better than 3.0 Å and R-factors less than 0.30 were then superposed onto the cyclic AMP dependent protein kinase (PKA) structure (PDB ID: 1STC, chain E) using the program Baton based on the method developed in Comparer (Sali et al, 1990). The details of their quality and the complex crystallised are listed in Appendix A.

### *3.2.3 Distance matrices construction*

The obtained structural alignment in format (.ali) was used to infer equivalent residues in the kinase superfamily using the program KinaseMap (Smith, 2006). Distances between every residue surrounding the pocket were measured from representative atoms in distinctive parts of the amino acids near the end of the side chains (see Table 4). Half-diagonal distance matrices were constructed for each PDB chain in the dataset.

### *3.2.4 Matrix correlation and dendrogram construction*

The correlations between the matrices were calculated by the Mantel test using program zt (Bonnet & Peer, 2002). The relationships between distance matrices were defined using the neighbour-joining algorithm from the program PHYLIP (Felsenstein, 2004). The sequences of the catalytic domains and also the residues in the active sites were aligned in order to calculate dendrograms for comparison with the shape-based dendrogram using program ClustalX. The dendrograms were made using the program TREEVIEW (Page, 1996).

### *3.2.5 Circular dendrogram construction*

The equivalent set of distances was also measured for a new set of 35 non-redundant crystal structures of kinases, which have been assayed by Fabian *et al.* (Fabian et al, 2005). These structures are the PDB files with the best resolutions that have the protein names matching those assayed by Fabian et al. and contain a variety of inhibitors bound in the ATP binding site. This allows the relationship between the spatial arrangement of residues of different kinases and their binding affinities ($K_d$) against 10 ligands to be visualised. The colour gradient representation of the tree plotting program iTOL (Letunic & Bork, 2007) was used where the intensity of the colours was proportional to $-\log_{10}K_d$ in order to allow comparison for millimolar to sub-nanomolar values of binding constants ($K_d$).

## 3.3 Results & Discussion

### *3.3.1 Residue selection for the construction of quasi-shape*

I wish to investigate whether the spatial arrangement of residues in the ATP binding pocket has an influence on which inhibitor the kinase recognises. In order to avoid comparing extremely variable regions of the pocket, I focused only on protein structures with staurosporine or adenine-ring containing compounds bound. A set of seven points was selected that can represent common features of the pocket. The Mantel test can distinguish the pockets of different kinases based on the assumption that the matrix of distances between points surrounding the adenosine pocket can reflect key features of the pocket shape in multiple dimensions; I call a matrix of this sort a "quasi-shape" (Figures 23 & 24).

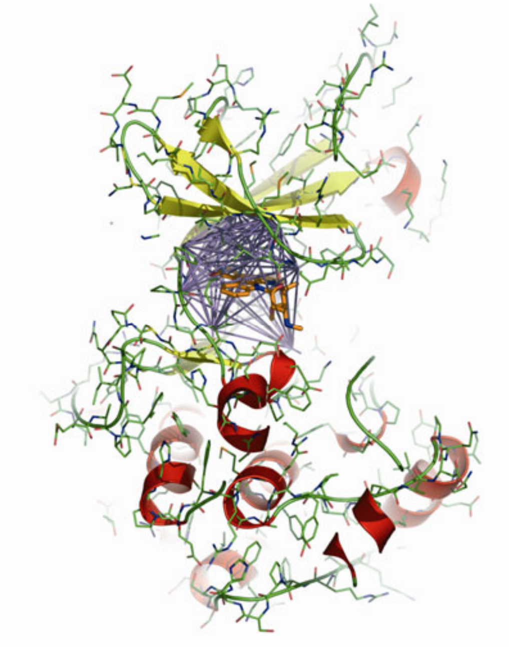


Figure 23. The quasi-shape is located between the N-terminal and the C-terminal lobes

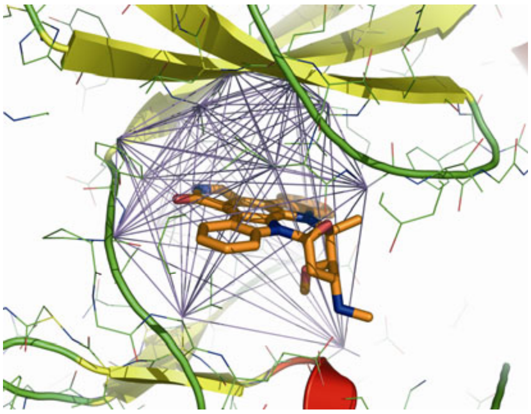
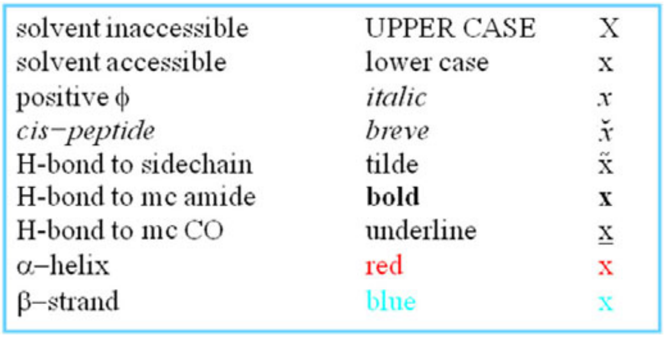


Figure 24. The quasi-shape (purple lines)

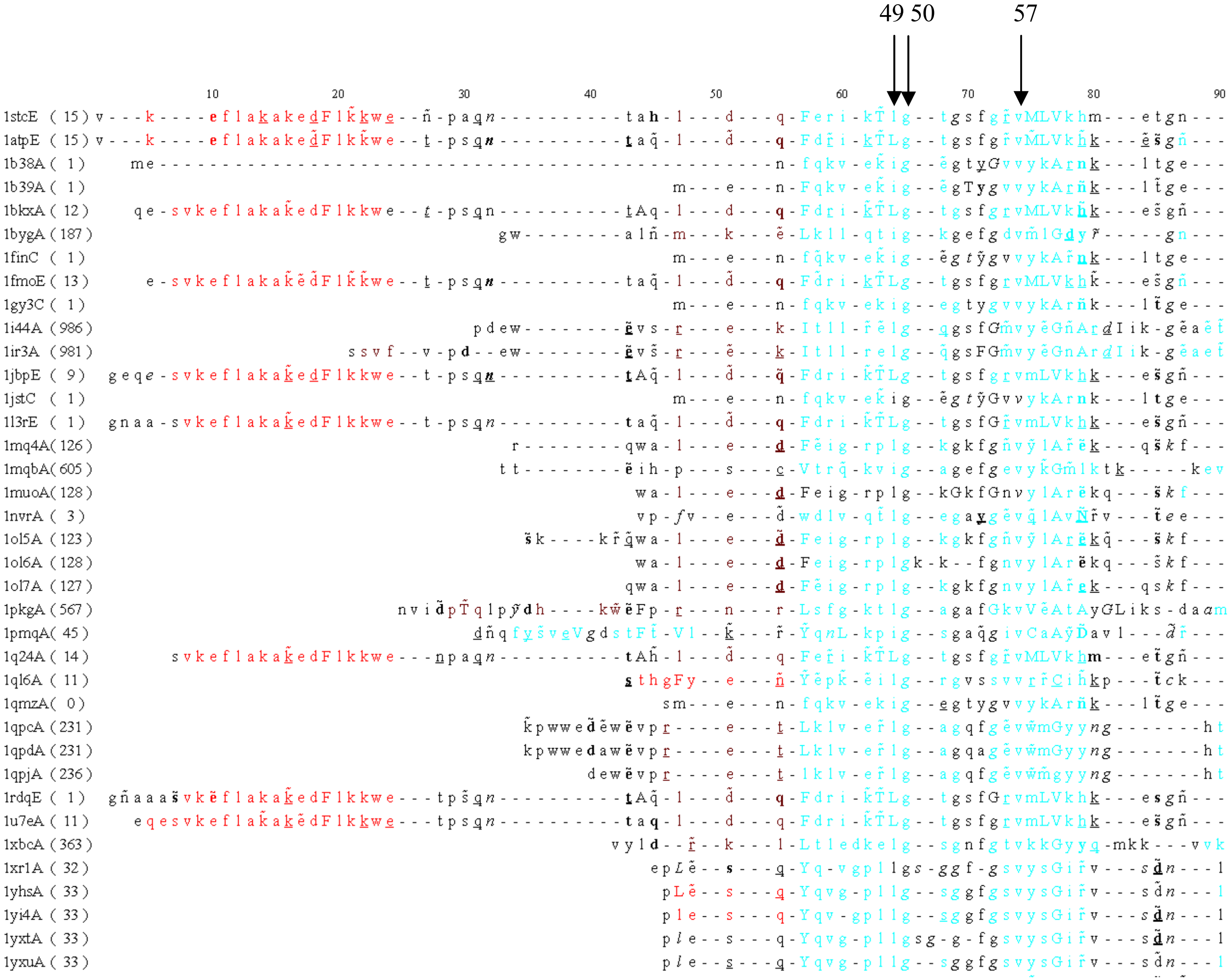
Calculated correlation coefficients among distance matrices of the same size and order of elements can be used to estimate the similarities in the spatial arrangements of side chains and hence the relationship between shape and the ability of the parent kinase to bind various inhibitors. The resulting shape-based dendrogram is constructed from 17 points in 17 residues, which are equivalent to the following residues in cAMP dependent protein kinase: Leu 49, Gly 50, Val 57, Ala70, Met 71, Lys 72, Val 104, Met 120, Glu 121, Tyr 122, Val 123, Glu 170, Asn 171, Thr 183, Asp 184, Glu 127, Leu 173.

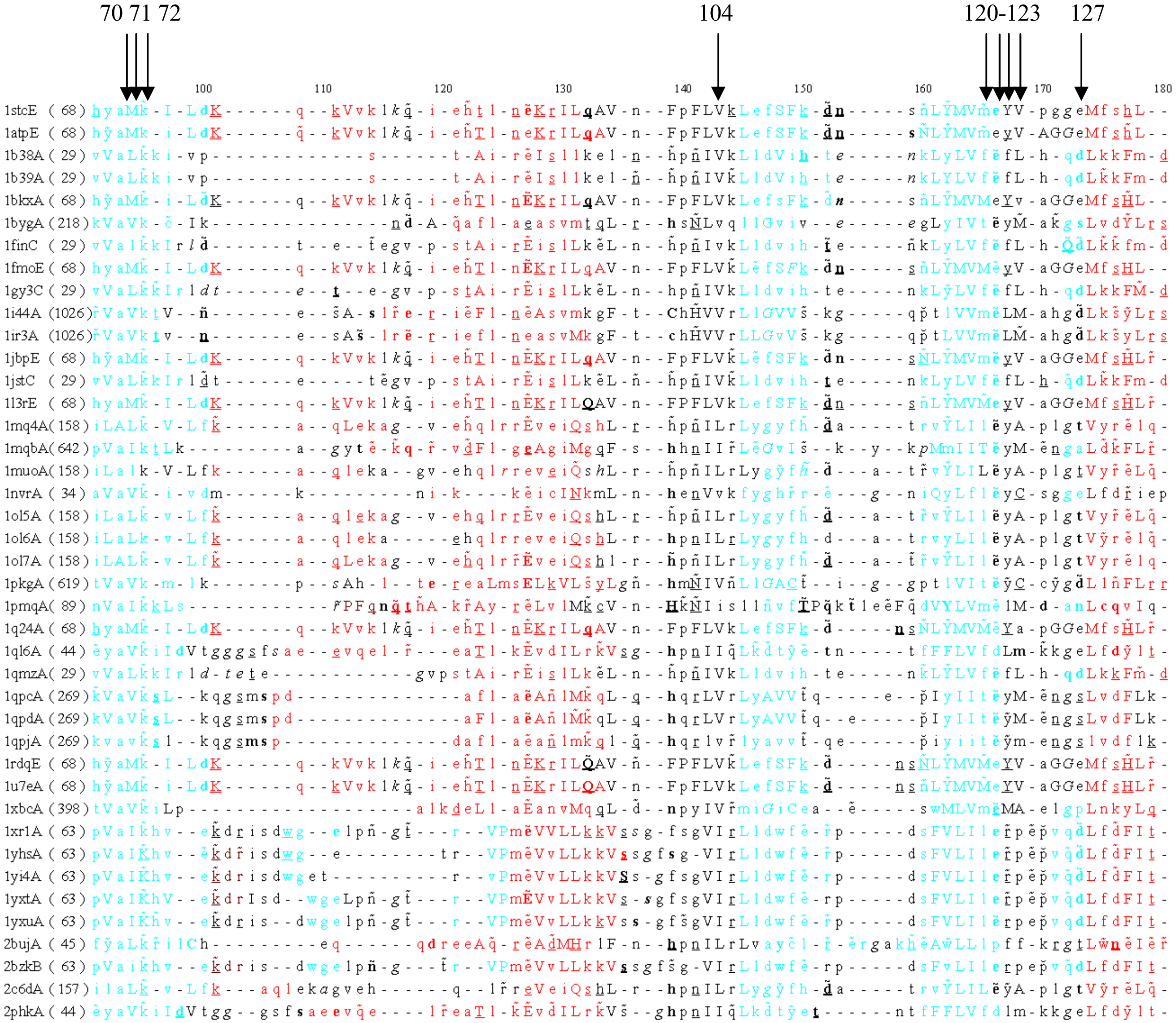
These residues can be found in the multiple structural alignment in the JOY format (Mizuguchi et al, 1998) in Figure 26.

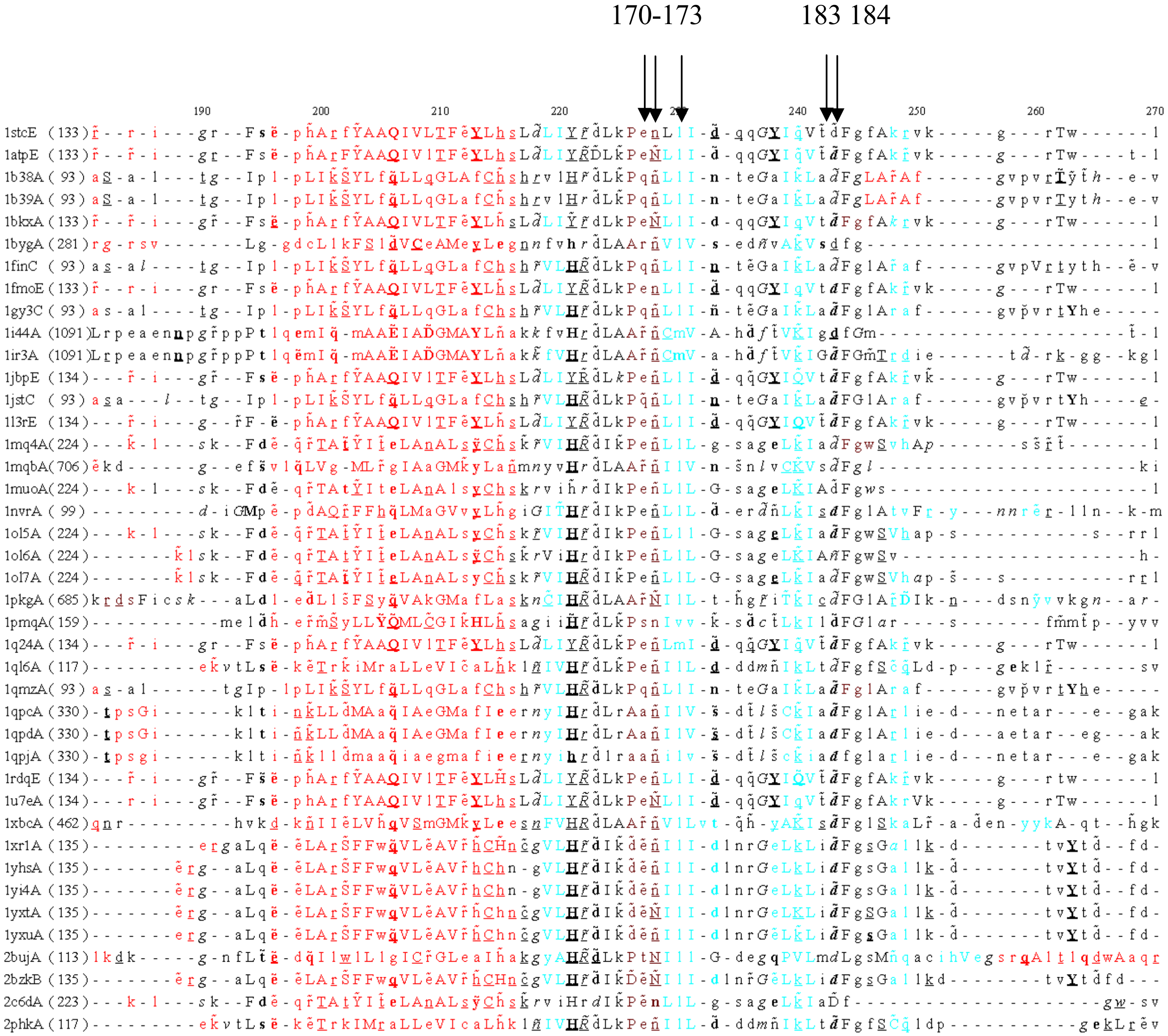


| | | |
|---|---|---|
| solvent inaccessible | UPPER CASE | X |
| solvent accessible | lower case | x |
| positive φ | *italic* | *x* |
| *cis–peptide* | *breve* | x̆ |
| H-bond to sidechain | tilde | x̃ |
| H-bond to mc amide | **bold** | **x** |
| H-bond to mc CO | underline | x |
| α–helix | red | x |
| β–strand | blue | x |

Figure 25. JOY annotations.


49 50
57
10
20
30
40
50
60
70
80
90
1stcE ( 15 )
1atpE ( 15 )
1b38A ( 1 )
1b39A ( 1 )
1bkxA ( 12 )
1bygA ( 187 )
1finC ( 1 )
1fmoE ( 13 )
1gy3C ( 1 )
1i44A ( 986 )
1ir3A ( 981 )
1jbpE ( 9 )
1jstC ( 1 )
1l3rE ( 1 )
1mq4A ( 126 )
1mqbA ( 605 )
1muoA ( 128 )
1nvrA ( 3 )
1ol5A ( 123 )
1ol6A ( 128 )
1ol7A ( 127 )
1pkgA ( 567 )
1pmqA ( 45 )
1q24A ( 14 )
1ql6A ( 11 )
1qmzA ( 0 )
1qpcA ( 231 )
1qpdA ( 231 )
1qpjA ( 236 )
1rdqE ( 1 )
1u7eA ( 11 )
1xbcA ( 363 )
1xr1A ( 32 )
1yhsA ( 33 )
1yi4A ( 33 )
1yxtA ( 33 )
1yxuA ( 33 )


70 71 72
104
120-123
127
100
110
120
130
140
150
160
170
180
1stcE ( 68)
1atpE ( 68)
1b38A ( 29)
1b39A ( 29)
1bkxA ( 68)
1bygA (218)
1finC ( 29)
1fmoE ( 68)
1gy3C ( 29)
1i44A (1026)
1ir3A (1026)
1jbpE ( 68)
1jstC ( 29)
1l3rE ( 68)
1mq4A(158)
1mqbA(642)
1muoA(158)
1nvrA ( 34)
1ol5A (158)
1ol6A (158)
1ol7A (158)
1pkgA (619)
1pmqA( 89)
1q24A ( 68)
1ql6A ( 44)
1qmzA( 29)
1qpcA (269)
1qpdA (269)
1qpjA (269)
1rdqE ( 68)
1u7eA ( 68)
1xbcA (398)
1xr1A ( 63)
1yhsA ( 63)
1yi4A ( 63)
1yxtA ( 63)
1yxuA ( 63)
2bujA ( 45)
2bzkB ( 63)
2c6dA (157)
2phkA ( 44)


170-173
183 184
190
200
210
220
240
250
260
270
1stcE (133)
1atpE (133)
1b38A (93)
1b39A (93)
1bkxA (133)
1bygA (281)
1finC (93)
1fmoE (133)
1gy3C (93)
1i44A (1091)
1ir3A (1091)
1jbpE (134)
1jstC (93)
1l3rE (134)
1mq4A (224)
1mqbA (706)
1muoA (224)
1nvrA (99)
1ol5A (224)
1ol6A (224)
1ol7A (224)
1pkgA (685)
1pmqA (159)
1q24A (134)
1ql6A (117)
1qmzA (93)
1qpcA (330)
1qpdA (330)
1qpjA (330)
1rdqE (134)
1u7eA (134)
1xbcA (462)
1xr1A (135)
1yhsA (135)
1yi4A (135)
1yxtA (135)
1yxuA (135)
2bujA (113)
2bzkB (135)
2c6dA (223)
2phkA (117)

```
                      280       290       300       310       320       330       340       350       360
1stcE  (199) c̃gtpeỹLAPẼIil-s---k--GỸn-k̃AVDWWaLGVLIYẼMAA-gypPFfad---qpi-qIyekIvs-G------------------kvr-
1atpE  (199) cGt̃peỹLAPẼIil-s---k--GỸñ-kAVDWWaLGVLIYẼMAA-gypPFfad---qpi-qIyekIvs-g------------------kvr-
1b38A  (164) V-tlwỸrAPẼiLL-gc--k-y-ỹs-taVD̃IW̃SLGC̃IFAẼMVt-rralFpGd̃---seid̃QLfr̃IFrtLgtP-dẽ--vvW-pgVt̃smpdykp
1b39A  (164) V-tlwỸrAPẼiLL-gc--k-y-ys-taVD̃IW̃SLGC̃IFAẼMVt-rralFpGd̃---seid̃QLfr̃IFrtLgtP-dẽ--vvW-pgVt̃smpdykp
1bkxA  (199) c̃gtpeỹLAPẼIil-s---k--GỸn-kAVD̃WWaLGVLIYẼMAA-gypPFfad---qpi-qIyekIvs-G------------------kvr-
1bygA  (348) l-pvkw̃tAPẽAlr̃-ek--k---fs-tkSDVWsFGILLW̃ẼIYsfGrvPYp-rI---plkd̃Vvpr̃Vẽk-----------------gy-kM̃d-
1finC  (164) V-tlwỸrAPẼILL-gc--k-y-Ys-taVDIW̃SLGC̃IFAẽMVt-rrAlFpGd̃---seid̃QLfr̃IFrtLgtP-dẽ--vvW-pgVt̃smpdykp
1fmoE  (199) c̃Gt̃peỸlAPẼIil-s---k--GỸñ-kAVDW̃WaLGVLIYẼMAA-gypPFfad---q̃pi-q̃IyekIvs-g------------------kvr-
1gy3C  (163) VVt̃lwỸrAPẼILL-gc̃--k-y-ỹs-taVDIW̃SLGC̃IFAẽMVt-rralFpGd̃---seid̃QLfr̃IFrtLgtP-dẽ--vvW-pgVt̃sm̃pdYkp
1i44A (1172) --pvr̃w̃mAPẽSLkdgv------ft-t̃sSDM̃WsFGVVLwẼIT̃slAeqPyq-gl--s̃ñ-eq̃VlkfVmdggyL-dq̃p-d--------------
1ir3A (1171) l-pvr̃w̃MAPẽSLkdgv------Ft-t̃sSDM̃WsFGVVLwẼIT̃slAeq̃Pyq-gl--s̃ñ-eq̃Vlk̃fVmd̃ggyl-dq̃p-d--------------
1jbpE  (199) c̃Gt̃peỸLAPẼIil-s---k--GỸñ-kAVDWWaLGVLIYẼMAA-gypPFfad---epi-qIyekIvs-g------------------kvr-
1jstC  (163) vVt̃lwỸrAPẼILL-gc--k-y-ỹs-taVDIW̃SLGC̃IFAẽMVt-rrAlFpGd̃---seid̃QLfr̃IFrtLgTP-dẽ--vvW-pgVt̃smpdỸkp
1l3rE  (199) c̃Gt̃peỸLAPẼIil-s---k--GỸn-kAVD̃WWaLGVLIYẼMAA-gypPFfad---qpi-qIyekIvs-g------------------kvr-
1mq4A  (290) C̃gtLd̃ỹlPPẼMIẽ-gr̃--m̃---H̃d-ẽkVD̃LWSLGVLCYẽFLV-gkpPFean---tyq-eTykrIs------------------rv-eft-
1mqbA  (780) --piR̃w̃tApẽAisyr---k---ft-saSDVWSFGIVMW̃ẼVMtyGerPYw-el--s̃ñh-ẽVmk̃aiñd̃-g------------------frLp
1muoA  (290) cgt̃lDỹlpPEmIe-g---r-m-hd-ekVDLWsLGVLCYẽFLv-gkpPFean---t̃ỹq̃-eTykrIsr̃vẽf-------------------t-
1nvrA  (168) c̃Gt̃lpYvAPẼLlk--r--r-efh̃A-epvD̃VWSCGIVLTaMLA-GẽlPW--dq̃P-s̃d̃s̃c̃q̃ẽỸsd̃Wk̃e------------------------
1ol5A  (290) c̃Gt̃ldỸLPPẽMiẽ-g---r-m-H̃d-ẽk̃VD̃LWsLGVLCYẽFLV-gkpPFeanty----qẽTykr̃Is̃r̃vẽf-------------------t-
1ol6A  (292) --t̃ldỹlPPẽmieg-m--h̃----d-ẽk̃VD̃LWSLGVLCyẽFLv-gkpPFeanty--q--etykrIsrv-------------------eft-
1ol7A  (290) cgt̃Ld̃ỹlPPẼmIẽ-g---r̃-m-H̃d-ẽkVD̃LWSLGVLCYẽFLV-gkpPFeant̃y----q̃etykrIs̃r̃vẽf-------------------t-
1pkgA  (831) l-pvkw̃MApẽS̃ifn------cvỸt-feS̃DVWsYGIFLw̃ẼLFsLGssPyp-gmpvds--kFykmIke-g---frm̃-l--------------
1pmqA  (226) --tryỸr̃APẼvIL-gm-----gỹk-eNVD̃IW̃SVGC̃IMGẽMVr̃-hkiLFpGr---dỹid̃QWnk̃VIeq̃lGtpc̃peFMkk̃Lqptvr̃nỹVẽ---
1q24A  (199) c̃gt̃peỸLAPẼIil-s---k--GỸn-kAVDWWaLGVLIYẼMAA-gypPFfad---qpi-qIyekIvs-g------------------kvr-
1ql6A  (184) cgtpsylAPẽIIẽC̃smndnh̃pGỸg-k̃eVD̃MWsTGVIMỹtLLA-gspPFwhr---kq̃m-lM̃lrmIm̃sgnyqfg--speW̃d-----------
1qmzA  (163) vVt̃lwỸrAPẼILL-gc--k-y-ỹs-taVDIW̃SLGC̃IFAẽMVt-rralFpGd---seid̃QLfr̃IFrtLgTP-dẽ--vvW-pgVt̃smpdYkp
1qpcA  (402) f-pikw̃tApẽAinyg------tFt-ikS̃DVWsFGILLT̃ẼIVthGriPYp---gmt̃n-pẽViqnler-gyr̃M̃vr̃pd---------------
1qpdA  (402) f-pikw̃tApẽAinyg------tFt-ikS̃DVWsFGILLT̃ẼIVthGriPYp---gmt̃n-pẽViqnler-gyr̃M̃vr̃pd---------------
1qpjA  (402) f-pikw̃tapẽainyg------tft-iks̃dvwsfgillt̃ẽivthgriPyp---gmt̃n-pẽviqnler-gyrm̃vr̃Pd---------------
1rdqE  (199) c̃Gt̃peaLAPẼIil-s-----kGỸn-kAVD̃WWaLGVLIYẽMAA-gypPFfad---qpi-qIyekIv------------------sG-kvr-
1u7eA  (199) agt̃peỸLAPẽIil-s-----kGỸn-k̃AVDWWaLGVLIYẼMAA-gypPFfad---qpi-qIyekIvs-G------------------kvr-
1xbcA  (534) w-pvkw̃YApẽC̃iny-----ỹ-kFs̃-skS̃DVWSFGVLMwẼAFsyGqkPYr---gmkgs-eVtamlek̃-gẽrmgc̃pa---------------
1xr1A  (203) -gt̃r̃vYSPPẽwIryh------rỸhGr̃saaVWsLGILLỸD̃MVC̃-gdiPFe-----h̃d̃e-ẽIir-----GqVfFr̃q----------------
1yhsA  (203) -Gt̃r̃vYSPPẽwIryh------rỸhGr̃saaVWsLGILLỸD̃MVC̃-gdiPFe-----hd̃e-eIir̃-----Gq̃VfFr̃q----------------
1yi4A  (203) -Gt̃r̃vySPPẽwIryh------rỸhGr̃saaVWsLGILLỸD̃MVC̃-gdiPFe-----h̃d̃e-ẽIir-----GqVfFr̃q----------------
1yxtA  (203) -Gt̃r̃vySPPẽwIryh------rỸhGr̃saaVWsLGILLỸD̃MVC-gdiPFe-----hd̃e-eIir-----GqVfFr̃q----------------
1yxuA  (203) -Gt̃r̃vySPPẽwIryh------rỸhGr̃saaVWsLGILLỸD̃MVC̃-gdiPFe-----hd̃e-eIir-----GqVfFr̃q----------------
2bujA  (194) C-t̃isyrAPẽLf--sVqshcv-Id̃-ẽr̃TD̃VW̃SLGC̃VLyAMMf-GeGPy----------dmvf--qk-g----------------------
2bzkB  (203) -Gt̃r̃vỹSPPẽwIryh------rỸhGrsaaVWsLGILLỸD̃MVC̃-gdiPFe-----hd̃e-eIig-----GqVfFr̃q----------------
2c6dA  (279) hGtldỹlPPẽmiẽ-g---r̃mh--d-ekVD̃LWSLGVLCYẼFLV-gkpPFeanty----qẽTyk̃r̃Is̃rv-eft-fpd---------------
2phkA  (184) c̃gt̃psylAPẽIiẽCsmndnh̃pGỸg-k̃eVD̃MWsTGVIMytLLa-gspPFwhr---kqm-lMlrm̃Ims-gñyq-fgspeWd-----------
```

```
                     370       380       390       400       410       420       430       440       450
1stcE  (257) -Fps-------------hFs̃----------sdLkd̃LLrñLLqvdltkR̃fGnlk̃ng-vnd̃Ik̃n--h̃kW̃Fa-ttd---wiaIyqr̃k-ve-Ap
1atpE  (257) -fps-------------hFs̃----------sdLkd̃LLrñLLqvd̃lt̃kR̃fGñlkng-vndIk̃n--h̃kW̃Fa-ttd---wiaIyq̃r̃k̃-ve-Ap
1b38A  (239) sFpkwarqdfskvVp--pLd----------ẽdGr̃s̃LLsqMLhYdpnkR̃i-----s-AkaAla--h̃pFFq--------------d̃-vt-kp
1b39A  (239) sFpkwarqdfskvVp--pLd----------ẽdGr̃s̃LLsqMLhYdpnkR̃i-----s-AkaAla--h̃pFFq--------------d̃-vt-kp
1bkxA  (257) -FPs-------------h̃Fs̃----------sdLkd̃LLrnLLqvd̃lt̃k̃R̃fGnlkñg-vnd̃Ik̃n--H̃kW̃Fa-ttd---wiaIyqrk-ve-ap
1bygA  (407) -aPd----------g---Cp----------paVyeVMkñCWhld̃aam̃R̃p-----s̃-Flq̃Lr̃----eq̃Lẽ---------hIk̃the l
1finC  (239) sFpkwarqd̃Fs̃kvVp--pLd----------ẽdGr̃s̃LLsq̃MLhYd̃pnk̃R̃i-----s-AkaAla--h̃pFFq--------------d̃-vt-k̃p
1fmoE  (257) -fps-------------hFs̃----------sdLkd̃LLrñLLqvd̃lt̃k̃R̃fGñlkng-vndIk̃n--h̃kW̃Fa-tTd---wiaIyqr̃k-ve-ap
1gy3C  (239) sFpkwarqd̃fs̃kvVp--pLd----------edGrs̃LLsq̃MLhYdpnkR̃i-----s-AkaAla--h̃pFFq--------------d̃-vt-kp
1i44A (1233) -----------------nC̃p----------er̃Vt̃dLM̃rmC̃WqfnpkmR̃PtFl--eIVñlLk̃dd̃Lh̃ps̃Fp-evS---ffh-seenk
1ir3A (1233) -----------------nC̃p----------erVt̃dLM̃rmC̃WqfnpkmR̃PtFl--eIVñlLk̃ddLh̃ps̃Fp-evS̃---Ffh̃-seenk
1jbpE  (257) -fps-------------h̃Fs̃----------sdLkd̃LLrnLLqvdltkR̃fGñlkñg-vnd̃Ik̃n--H̃kW̃Fa-ttd---wiaIyqr̃k-ve-Ap
1jstC  (239) sFpk̃warqd̃fsk̃VVp--pLd----------edGrs̃LLsq̃MLhYd̃pnk̃R̃i-----s-AkaAla--h̃pFFq--------------d̃-vt-kp
1l3rE  (257) -fps-------------h̃Fs̃----------sd̃Lkd̃LLrnLLqvd̃lt̃kR̃fGnlkng-vndIk̃n--h̃k̃W̃Fa-ttd---wiaIyqr̃k-ve-Ap
1mq4A  (348) -fpd-------------fVt̃----------ẽgAr̃d̃LIsr̃LLkhnpsqR̃p------mLreVle--hpW̃It-añs̃s̃
1mqbA  (838) -tPm-------------dCP----------saIỹqLMm̃qCWq̃q̃erar̃R̃p-----k-FadIvsiLd̃kLir---------apdsLk--t---
1muoA  (348) -fpd-------------fVt----------egArd̃LIsrLLkhñps̃qrp--m----lreVle--H̃pW̃It-anss
1nvrA  (224) ------kk̃tylnP̌Wk--k̃Id----------sAPlaLLhk̃ILvẽnpsaR̃i-----t-ipd̃Ikk̃--D̃rW̃ynkplk̃k---------------
1ol5A  (348) -fpd-------------fVt̃----------egArd̃LIsr̃LLkh̃ñps̃qR̃p------m̃LreVle--h̃pW̃It-añs̃s
1ol6A  (348) -fpd-------------fVt̃----------egArd̃LIsr̃LLkhnpsqR̃P------m̃lrẽVle--h̃pW̃It-ans̃s
1ol7A  (348) -fpd-------------fVt̃----------egArd̃LIsr̃LLkhñps̃qR̃p------mlreVle--h̃pW̃It-añS̃s̃--kp
1pkgA  (891) -sPe-------------hAp----------aeM̃ydIMktCWd̃adplk̃R̃p-----t-FkqIvq----lIe-kq̃
1pmqA  (300) nr̃pkyaGltfpkLFpd̃slFpad̃sehñklk̃asq̃Ar̃d̃LLskMLvID̃pak̃R̃i-----s̃-Vd̃d̃ALq--h̃pỸInvw̃ỹdp---------aẽVẽ-ap
1q24A  (257) -fps-------------hFs̃----------sdLkd̃LLrnLLqvdltkR̃fGnlkng-vndIk̃n--H̃kW̃Fa-ttd---wiaIyqr̃k-ve-Ap
1ql6A  (255) -----------------dys̃----------dtVkd̃LVsr̃FLvvq̃pq̃k̃R̃y-----t-AẽeAla--H̃pFFqqy------v
1qmzA  (239) sFpkwarqd̃fs̃kvVp--pLd----------edGrs̃LLsq̃MLhYdpnkR̃i-----s-AkaAla--h̃pFFq--------------d̃-vt-kp
1qpcA  (464) ---------------------------nCpẽeLỸq̃LMr̃lCWkẽr̃pẽd̃R̃p-----t-Fd̃yLr̃svLẽdfftate
1qpdA  (464) ---------------------------nCpẽeLỸq̃LMr̃lCWkẽr̃ped̃R̃p-----t-Fd̃yLr̃svLẽdfftat̃e
1qpjA  (464) ---------------------------ncpẽelỹq̃lmr̃lcwker̃pẽd̃r̃p-----t-fd̃ylr̃svlẽdfftat̃
1rdqE  (257) -Fps-------------h̃Fs̃----------sdLkd̃LLrñLLqvd̃lt̃kR̃fGnlkng-vndIk̃n--h̃kW̃Fa-ttd---wiaIyq̃r̃k̃-ve-ap
1u7eA  (257) -Fps-------------h̃Fs̃----------sdLkd̃LLrnLLqvd̃lt̃k̃R̃fGnlkng-vndIk̃n--h̃kW̃Fa-ttd---wiaIyqr̃k-ve-Ap
1xbcA  (596) ---------------------------gCpr̃ẽMỹdLMñlCWtyd̃veñR̃p-----g-FaaVẼlrLr̃nỹỹydvvne
1xr1A  (258) rVs̃s---------------------------eCq̃h̃LIr̃wCLalrpsdR̃P-----t̃-feẽIq̃n--H̃pW̃Mqd-vll--pqeT̃AeiH̃Lhsl--
1yhsA  (258) ---------------------------rVs̃-eC̃q̃h̃LIr̃wCLalr̃psd̃R̃P-----t̃-feẽIq̃n--H̃pW̃Mqd-vll--pq̃eT̃Aẽ-ih̃Lh
1yi4A  (258) ---------------------------rVs̃-eC̃q̃h̃LIr̃wCLalrpsd̃R̃P-----t̃-feẽIq̃n--H̃pW̃Mqd-vll--pqeT̃Ae-ih̃Lh
1yxtA  (258) rVs̃s---------------------------eCq̃hLIr̃wCLalrpsd̃R̃P-----t̃-feẽIq̃n--HpW̃Mqd-vll--pqeT̃Ae-ih̃Lhs
1yxuA  (258) rVs̃s---------------------------eC̃q̃h̃LIr̃wCLalr̃psd̃R̃P-----t-feeIq̃n--HpW̃Mqd-vll--pqeT̃Ae-ih̃Lh
2bujA  (243) -------dsValaVqnq-ipq̃sprHs̃----saLw̃qLLnsMmtvdphqR̃p-----h̃-IplLLs---qLẽalqPpapg
2bzkB  (258) rVs̃s---------------------------eC̃q̃h̃LIr̃wCLalr̃psd̃R̃P-----t̃-feẽIq̃n--H̃pW̃Mqd-vll--pq̃eT̃Aẽ-ih̃Lh
2c6dA  (350) -----------------fVt̃----------egArd̃LIsr̃LLkhnpsqR̃p------mLreVle--h̃pW̃It-añS̃skp
2phkA  (255) ---------------------------dys̃dtVkd̃LVsr̃FLvvq̃pq̃kR̃y-----t̃-AeẽAla--H̃pFFqq----y
```

```
                                 460             470             480             490
1stcE  (314) f i p k f k - g p G d̃ t s̃ ñ f d d y ẽ ẽ ẽ e i r v - i n e k c g k e F s e F
1atpE  (314) f i P k f k - g p g d̃ t s̃ n f d d y ẽ ẽ ẽ ẽ i r v s i n e k̃ c g k e F t e F
1b38A  (293) - v P h L r l
1b39A  (293) - v P h L r l
1bkxA  (314) f i P k f k - g p G d̃ t s̃ n f d d̃ y e ẽ e e i r v - i n e k̃ c G k e F t e F
1bygA
1finC  (293) - v p h L r̃ l
1fmoE  (314) f i P k F k - g p G d̃ t s̃ n f d d y e ẽ e e i r v - i n e k c g k e F t e F
1gy3C  (293) - v P h l
1i44A
1ir3A
1jbpE  (314) f i P k f k - g p G d̃ t s̃ n f d d̃ y e ẽ e e i r v - i n e k c g k e F t e F
1jstC  (293) - v P h L r̃ l
1l3rE  (314) f i P k f k - g p g d̃ t s̃ n f d d̃ y e ẽ ẽ e i r v - i n e k c g k e F t e F
1mq4A
1mqbA  (884) - l a d - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - f
1muoA
1nvrA  (272) - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - g a
1ol5A
1ol6A
1ol7A
1pkgA
1pmqA  (372) - p p q - - - - - - - - - q l d e r e h̃ t̃ i - - - - - - ẽ ẽ W̃ k ẽ l I ỹ k̃ ẽ V m̃ ñ
1q24A  (314) f i P k f k - g p g d̃ t s̃ n f d d̃ y e ẽ ẽ ẽ i r v - i n e k c g k e F s e F
1ql6A
1qmzA  (293) - v P h l
1qpcA
1qpdA
1qpjA
1rdqE  (314) f i P k f k - g p g d̃ t s̃ n f d d̃ y e ẽ e e i r v - i n e k c g k e F t e F
1u7eA  (314) f i P k f k - g p g d̃ t s̃ n f d d̃ y e ẽ e e i r v - i n e k c g k e F t e f
```

Figure 26. Multiple alignments from superposition of the kinases using program Baton.

### *3.3.2 Dendrogram constructed from the quasi-shape*

The result suggested that Mantel test correlations between the matrices derived from a small set of inter-atomic distances, between 7 frequently occurring atoms, can separate the majority of staurosporine complexes from the adenine containing complexes (Figure 27). This means there are observable differences in spatial arrangement of these atoms when staurosporine and adenine are bound. Thus, the Mantel Test appears to work well for classifying different three dimensional geometric shapes. However, the same kinase in different crystal forms can be scattered throughout the resulting shape-based dendrogram, implying that similarities between these conserved atoms are not sufficient to differentiate the structures of kinases (Figure 27).

Therefore, I investigated the use of distinctive parts near the ends of each amino acid residue as the centres for distance measurements in the construction of the distance matrices, thus allowing the derivation of a quasi-shape from each PDB file. The choice of these atoms, which depends on their residue type, is shown in Figure 11. By gradually increasing the number of residue points, I was able to categorise the same kinase in different crystal forms and different complexes into the same branch of the dendrogram. This dendrogram places the same type of kinase in different complexes in the same branch regardless of the bound ligand, and the staurosporine binding structures were clustered into one half of the tree (Figure 28). Therefore, the constructed matrices appear to be able to represent similar pockets.

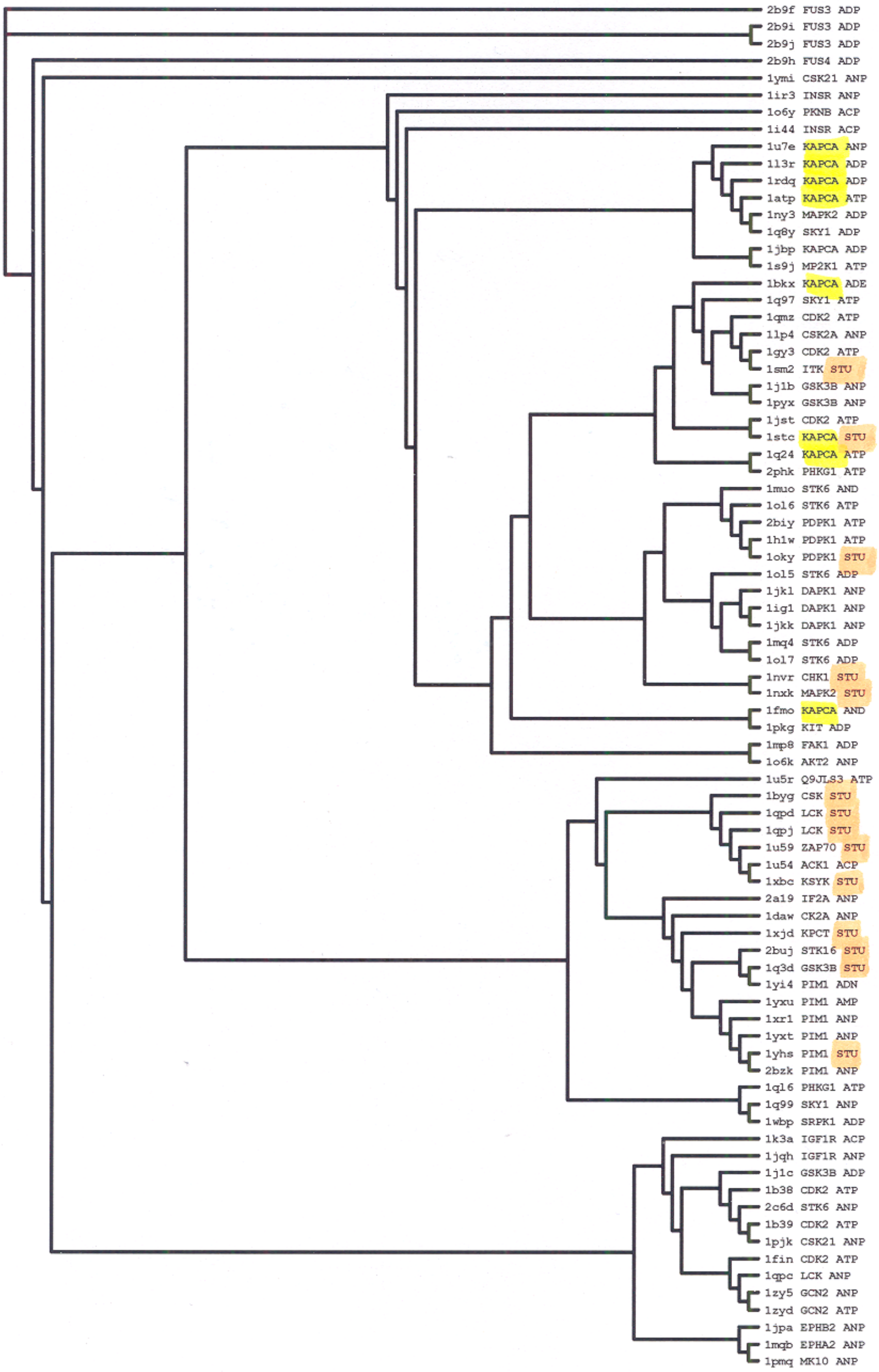


Figure 27. Mantel test can distinguish the shape derived from 7 frequently occurring atoms in staurosporine complex from adenosine phosphate complex, but structures of the same kinase, e.g. KAPCA are scattered throughout the tree.

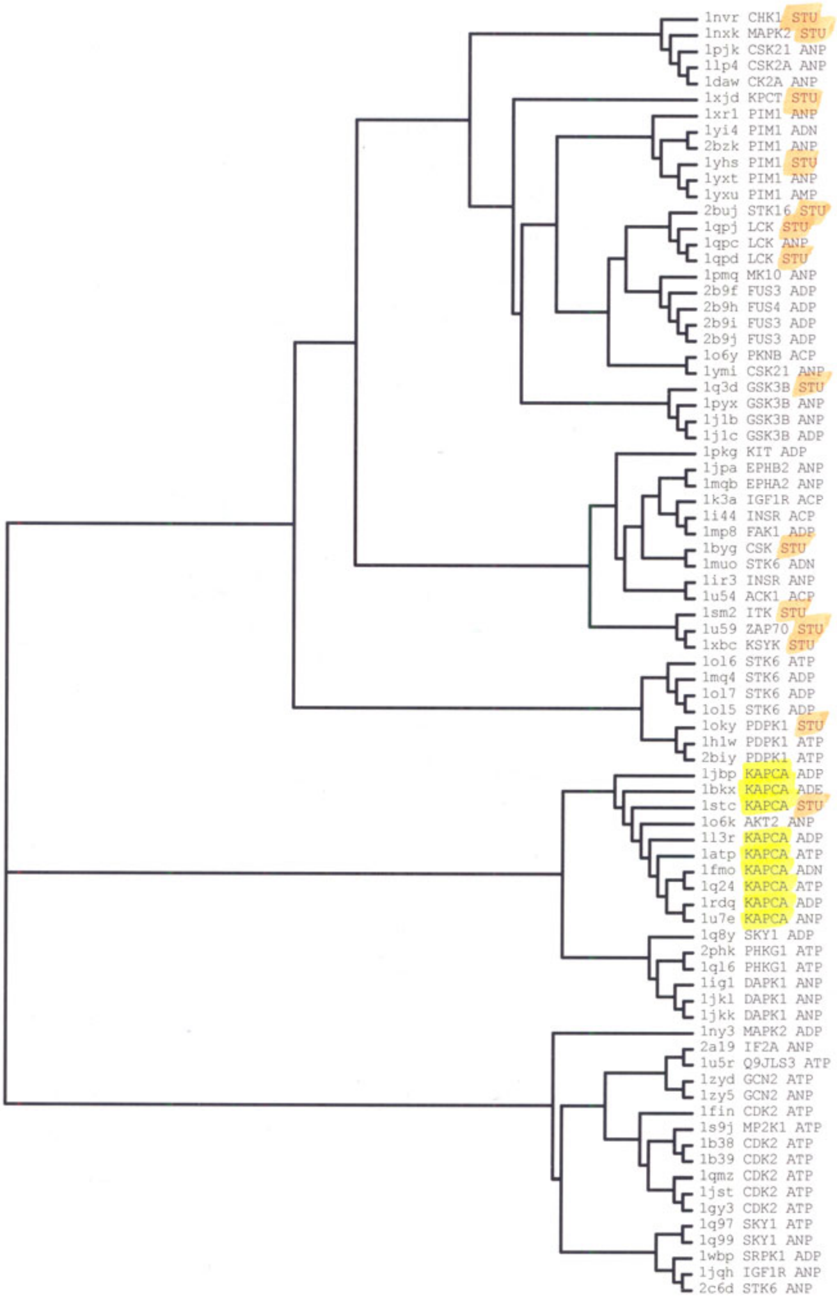
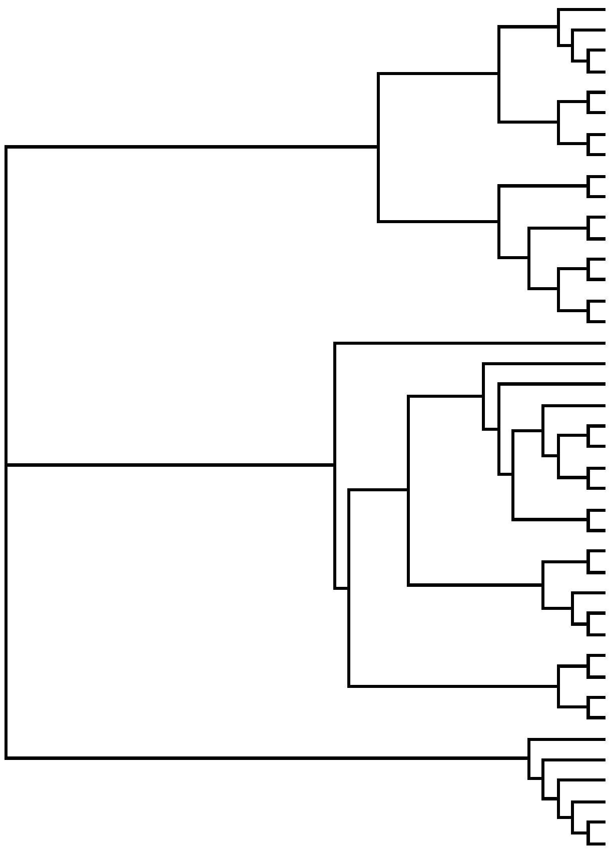


Figure 28. The shape-based dendogram constructed from 17 points in 17 residues, which places the same type of kinase in different complexes in the same branch regardless of the bound ligand.

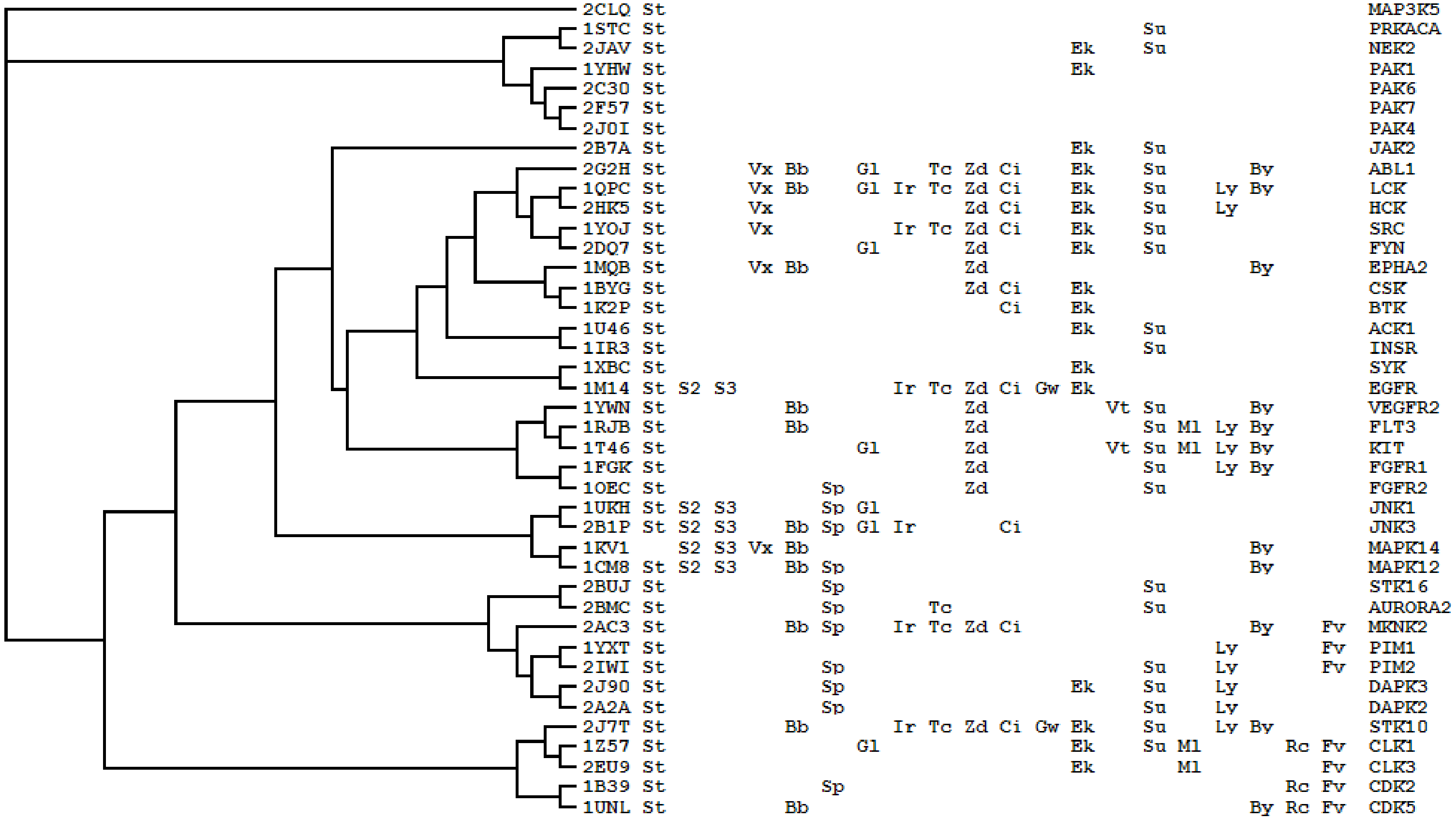
### *3.3.3 Sequence relationship with inhibitor binding*

I next investigated the relationship between the similarity obtained from sequence alignment and the ligand selectivity. However, it is difficult to conclude from these dendrograms whether the whole domain sequence alignment (Figure 29), the ligand accessible region alignment (Figure 30), or the shape similarity (Figure 31) dendrogram is better because these dendrograms cannot represent that very weak binding is somewhat similar to not binding at all. Hence, I also produced circular dendrograms using gradient colour to represent the relationship between sequence and shape to the binding affinities to 10 inhibitors so that the binding can be compared by the intensity of the colour.



| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1STC | St | | | | | | | | | | | | | | Su | | | | | | PRKACA |
| 2AC3 | St | | | | Bb | Sp | | Ir | Tc | Zd | Ci | | | | | | | By | | Fv | MKNK2 |
| 2J90 | St | | | | | Sp | | | | | | | Ek | | Su | | Ly | | | | DAPK3 |
| 2A2A | St | | | | | Sp | | | | | | | | | Su | | Ly | | | | DAPK2 |
| 2JAV | St | | | | | | | | | | | | Ek | | Su | | | | | | NEK2 |
| 2BMC | St | | | | | Sp | | | Tc | | | | | | Su | | | | | | AURORA2 |
| 1YXT | St | | | | | | | | | | | | | | | | Ly | | | Fv | PIM1 |
| 2IWI | St | | | | | Sp | | | | | | | | | Su | | Ly | | | Fv | PIM2 |
| 1Z57 | St | | | | | | Gl | | | | | | Ek | | Su | Ml | | | Rc | Fv | CLK1 |
| 2EU9 | St | | | | | | | | | | | | Ek | | | Ml | | | | Fv | CLK3 |
| 1B39 | St | | | | | Sp | | | | | | | | | | | | | Rc | Fv | CDK2 |
| 1UNL | St | | | | Bb | | | | | | | | | | | | | By | Rc | Fv | CDK5 |
| 1UKH | St | S2 | S3 | | | Sp | Gl | | | | | | | | | | | | | | JNK1 |
| 2B1P | St | S2 | S3 | | Bb | Sp | Gl | Ir | | | Ci | | | | | | | | | | JNK3 |
| 1KV1 | | S2 | S3 | Vx | Bb | | | | | | | | | | | | | By | | | MAPK14 |
| 1CM8 | St | S2 | S3 | | Bb | Sp | | | | | | | | | | | | By | | | MAPK12 |
| 2BUJ | St | | | | | Sp | | | | | | | | | Su | | | | | | STK16 |
| 1IR3 | St | | | | | | | | | | | | | | Su | | | | | | INSR |
| 1MQB | St | | | Vx | Bb | | | | | Zd | | | | | | | | By | | | EPHA2 |
| 2G2H | St | | | Vx | Bb | | Gl | | Tc | Zd | Ci | | Ek | | Su | | | By | | | ABL1 |
| 1QPC | St | | | Vx | Bb | | Gl | Ir | Tc | Zd | Ci | | Ek | | Su | | Ly | By | | | LCK |
| 2HK5 | St | | | Vx | | | | | | Zd | Ci | | Ek | | Su | | Ly | | | | HCK |
| 1YOJ | St | | | Vx | | | | Ir | Tc | Zd | Ci | | Ek | | Su | | | | | | SRC |
| 2DQ7 | St | | | | | | Gl | | | Zd | | | Ek | | Su | | | | | | FYN |
| 1BYG | St | | | | | | | | | Zd | Ci | | Ek | | | | | | | | CSK |
| 1K2P | St | | | | | | | | | | Ci | | Ek | | | | | | | | BTK |
| 1RJB | St | | | | Bb | | | | | Zd | | | | | Su | Ml | Ly | By | | | FLT3 |
| 1T46 | St | | | | | | Gl | | | Zd | | | | Vt | Su | Ml | Ly | By | | | KIT |
| 1YWN | St | | | | Bb | | | | | Zd | | | | Vt | Su | | | By | | | VEGFR2 |
| 1FGK | St | | | | | | | | | Zd | | | | | Su | | Ly | By | | | FGFR1 |
| 1OEC | St | | | | | Sp | | | | Zd | | | | | Su | | | | | | FGFR2 |
| 2B7A | St | | | | | | | | | | | | Ek | | Su | | | | | | JAK2 |
| 1U46 | St | | | | | | | | | | | | Ek | | Su | | | | | | ACK1 |
| 1XBC | St | | | | | | | | | | | | Ek | | | | | | | | SYK |
| 1M14 | St | S2 | S3 | | | | | Ir | Tc | Zd | Ci | Gw | Ek | | | | | | | | EGFR |
| 2CLQ | St | | | | | | | | | | | | | | | | | | | | MAP3K5 |
| 2J7T | St | | | | Bb | | | Ir | Tc | Zd | Ci | Gw | Ek | | Su | | Ly | By | | | STK10 |
| 1YHW | St | | | | | | | | | | | | Ek | | | | | | | | PAK1 |
| 2C30 | St | | | | | | | | | | | | | | | | | | | | PAK6 |
| 2F57 | St | | | | | | | | | | | | | | | | | | | | PAK7 |
| 2J0I | St | | | | | | | | | | | | | | | | | | | | PAK4 |

Figure 29. Phylogenetic tree obtained from neighbour joining clustering of Baton's structural sequence alignment of the whole kinase domain. The first column is the PDB ID, the next twenty columns are inhibitors from supplementary information 4 of Fabian et al., and the final column is the kinase name.



| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2CLQ | St | | | | | | | | | | | | | | | | | | | | MAP3K5 |
| 1STC | St | | | | | | | | | | | | | | Su | | | | | | PRKACA |
| 2JAV | St | | | | | | | | | | | | Ek | | Su | | | | | | NEK2 |
| 1YHW | St | | | | | | | | | | | | Ek | | | | | | | | PAK1 |
| 2C30 | St | | | | | | | | | | | | | | | | | | | | PAK6 |
| 2F57 | St | | | | | | | | | | | | | | | | | | | | PAK7 |
| 2J0I | St | | | | | | | | | | | | | | | | | | | | PAK4 |
| 2B7A | St | | | | | | | | | | | | Ek | | Su | | | | | | JAK2 |
| 2G2H | St | | | Vx | Bb | | Gl | | Tc | Zd | Ci | | Ek | | Su | | | By | | | ABL1 |
| 1QPC | St | | | Vx | Bb | | Gl | Ir | Tc | Zd | Ci | | Ek | | Su | | Ly | By | | | LCK |
| 2HK5 | St | | | Vx | | | | | | Zd | Ci | | Ek | | Su | | Ly | | | | HCK |
| 1YOJ | St | | | Vx | | | | Ir | Tc | Zd | Ci | | Ek | | Su | | | | | | SRC |
| 2DQ7 | St | | | | | | Gl | | | Zd | | | Ek | | Su | | | | | | FYN |
| 1MQB | St | | | Vx | Bb | | | | | Zd | | | | | | | | By | | | EPHA2 |
| 1BYG | St | | | | | | | | | Zd | Ci | | Ek | | | | | | | | CSK |
| 1K2P | St | | | | | | | | | | Ci | | Ek | | | | | | | | BTK |
| 1U46 | St | | | | | | | | | | | | Ek | | Su | | | | | | ACK1 |
| 1IR3 | St | | | | | | | | | | | | | | Su | | | | | | INSR |
| 1XBC | St | | | | | | | | | | | | Ek | | | | | | | | SYK |
| 1M14 | St | S2 | S3 | | | | | Ir | Tc | Zd | Ci | Gw | Ek | | | | | | | | EGFR |
| 1YWN | St | | | | Bb | | | | | Zd | | | | Vt | Su | | | By | | | VEGFR2 |
| 1RJB | St | | | | Bb | | | | | Zd | | | | | Su | Ml | Ly | By | | | FLT3 |
| 1T46 | St | | | | | | Gl | | | Zd | | | | Vt | Su | Ml | Ly | By | | | KIT |
| 1FGK | St | | | | | | | | | Zd | | | | | Su | | Ly | By | | | FGFR1 |
| 1OEC | St | | | | | Sp | | | | Zd | | | | | Su | | | | | | FGFR2 |
| 1UKH | St | S2 | S3 | | | Sp | Gl | | | | | | | | | | | | | | JNK1 |
| 2B1P | St | S2 | S3 | | Bb | Sp | Gl | Ir | | | Ci | | | | | | | | | | JNK3 |
| 1KV1 | | S2 | S3 | Vx | Bb | | | | | | | | | | | | | By | | | MAPK14 |
| 1CM8 | St | S2 | S3 | | Bb | Sp | | | | | | | | | | | | By | | | MAPK12 |
| 2BUJ | St | | | | | Sp | | | | | | | | | Su | | | | | | STK16 |
| 2BMC | St | | | | | Sp | | | Tc | | | | | | Su | | | | | | AURORA2 |
| 2AC3 | St | | | | Bb | Sp | | Ir | Tc | Zd | Ci | | | | | | | By | | Fv | MKNK2 |
| 1YXT | St | | | | | | | | | | | | | | | | Ly | | | Fv | PIM1 |
| 2IWI | St | | | | | Sp | | | | | | | | | Su | | Ly | | | Fv | PIM2 |
| 2J90 | St | | | | | Sp | | | | | | | Ek | | Su | | Ly | | | | DAPK3 |
| 2A2A | St | | | | | Sp | | | | | | | | | Su | | Ly | | | | DAPK2 |
| 2J7T | St | | | | Bb | | | Ir | Tc | Zd | Ci | Gw | Ek | | Su | | Ly | By | | | STK10 |
| 1Z57 | St | | | | | | Gl | | | | | | Ek | | Su | Ml | | | Rc | Fv | CLK1 |
| 2EU9 | St | | | | | | | | | | | | Ek | | | Ml | | | | Fv | CLK3 |
| 1B39 | St | | | | | Sp | | | | | | | | | | | | | Rc | Fv | CDK2 |
| 1UNL | St | | | | Bb | | | | | | | | | | | | | By | Rc | Fv | CDK5 |

Figure 30. Phylogenetic tree obtained from neighbour joining clustering of Baton’s structural alignment of residues in the ligand accessible region (closer than 3Å from any part of the ligand from 14 non-redundant human kinases)

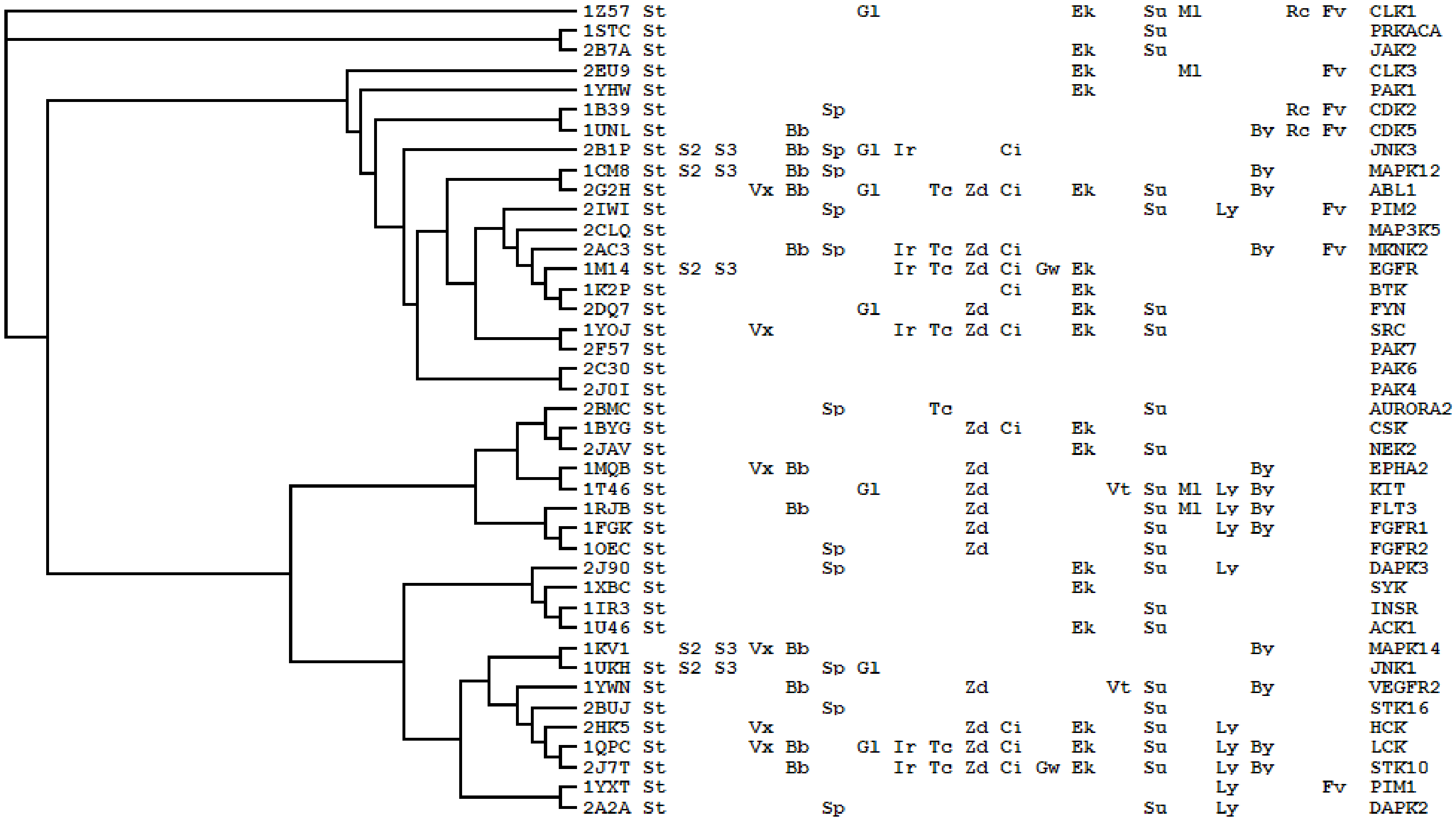

1Z57 St Gl Ek Su Ml Rc Fv CLK1
1STC St Su PRKACA
2B7A St Ek Su JAK2
2EU9 St Ek Ml Fv CLK3
1YHW St Ek PAK1
1B39 St Sp Rc Fv CDK2
1UNL St Bb By Rc Fv CDK5
2B1P St S2 S3 Bb Sp Gl Ir Ci JNK3
1CM8 St S2 S3 Bb Sp By MAPK12
2G2H St Vx Bb Gl Tc Zd Ci Ek Su By ABL1
2IWI St Sp Su Ly Fv PIM2
2CLQ St MAP3K5
2AC3 St Bb Sp Ir Tc Zd Ci By Fv MKNK2
1M14 St S2 S3 Ir Tc Zd Ci Gw Ek EGFR
1K2P St Ci Ek BTK
2DQ7 St Gl Zd Ek Su FYN
1YOJ St Vx Ir Tc Zd Ci Ek Su SRC
2F57 St PAK7
2C30 St PAK6
2J0I St PAK4
2BMC St Sp Tc Su AURORA2
1BYG St Zd Ci Ek CSK
2JAV St Ek Su NEK2
1MQB St Vx Bb Zd By EPHA2
1T46 St Gl Zd Vt Su Ml Ly By KIT
1RJB St Bb Zd Su Ml Ly By FLT3
1FGK St Zd Su Ly By FGFR1
1OEC St Sp Zd Su FGFR2
2J90 St Sp Ek Su Ly DAPK3
1XBC St Ek SYK
1IR3 St Su INSR
1U46 St Ek Su ACK1
1KV1 S2 S3 Vx Bb By MAPK14
1UKH St S2 S3 Sp Gl JNK1
1YWN St Bb Zd Vt Su By VEGFR2
2BUJ St Sp Su STK16
2HK5 St Vx Zd Ci Ek Su Ly HCK
1QPC St Vx Bb Gl Ir Tc Zd Ci Ek Su Ly By LCK
2J7T St Bb Ir Tc Zd Ci Gw Ek Su Ly By STK10
1YXT St Ly Fv PIM1
2A2A St Sp Su Ly DAPK2

Figure 31. Phylogenetic tree obtained from neighbour joining clustering of Mantel's correlation of distance matrices

### *3.3.4 Dendrogram displaying the relationship between shape and inhibitor binding affinities*

The general sequence-based dendrogram (Figure 32) is shown for comparison with the shape comparison dendrogram (Figure 33). It can be seen that kinases with similar sequence does not always bind similar inhibitors.

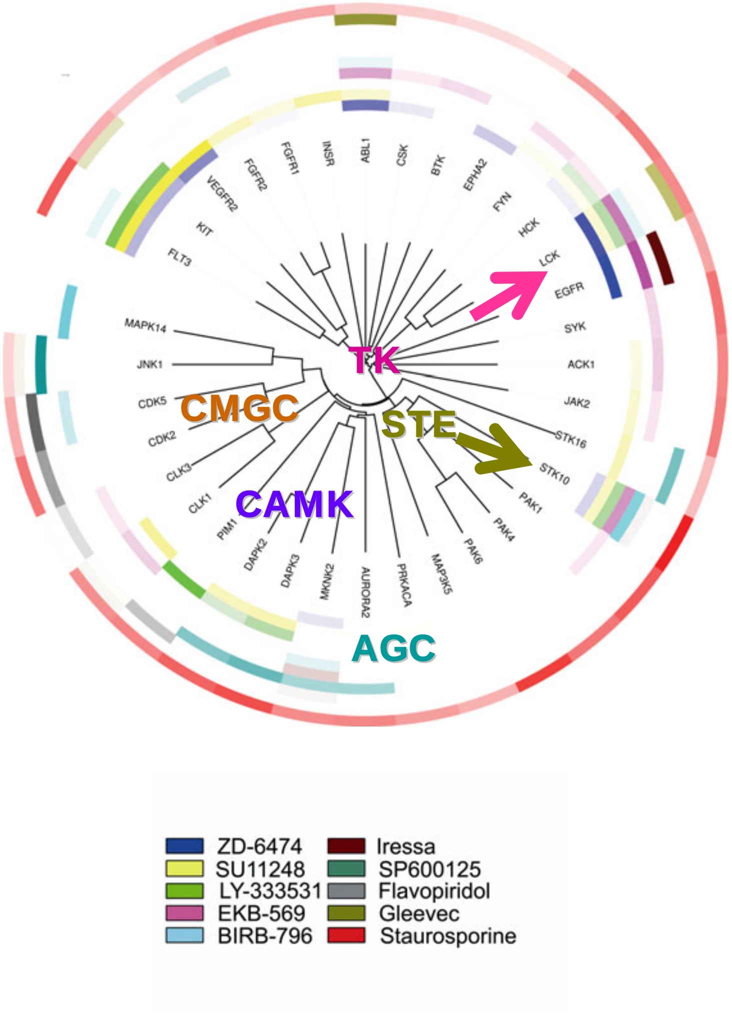


Figure 32. The classic dendrogram based on sequence similarity from structure based sequence alignment using the program Baton. The intensity of the colour is proportional to -log$_{10}$Kd of the inhibitor. The kinases with closest quasi-shape similarity are marked with arrows.

When I apply the same dendrogram colouring method to the set of 35 non-redundant structures where the $K_d$ of the enzyme has been studied by Fabian *et al.* (Fabian et al, 2005), I obtain a dendrogram that characterises the ability to bind ten ligands, based on the similarity in quasi-shape as shown below (Figure 33).

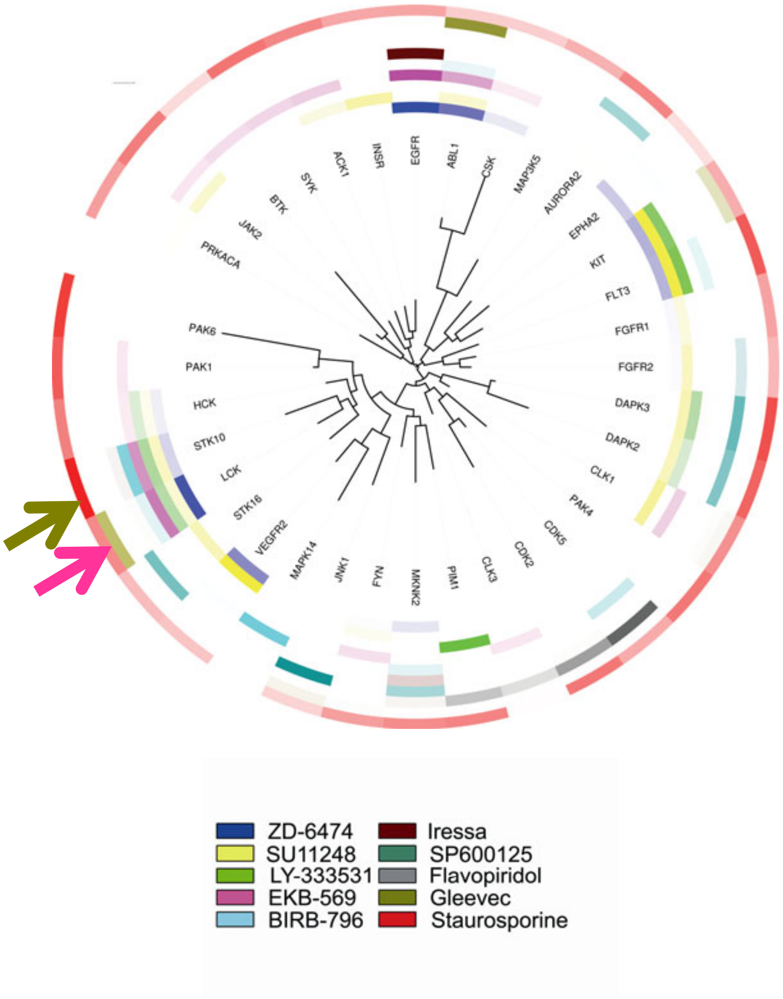


Figure 33. The shape-based dendrogram shows the matrix correlation between the shapes of the kinases and their binding affinities to 10 inhibitors. The kinases with closest quasi-shape similarity are marked with arrows.

It is evident that kinases with similar pocket quasi-shapes can have similar inhibitor binding profiles, regardless of their family membership. A nice example is serine/threonine kinase 10 (STK10), which is clustered in the sequence-based dendrogram (Figure 32) as an STE kinase as defined in the protein kinase phylogenetic tree by Manning *et al.* (Manning et al, 2002). When considering STK10 in terms of similarity in spatial arrangement of residues (Figure 33), it is instead paired with leukocyte-specific protein tyrosine kinase (LCK) which is a tyrosine kinase. The sequences are quite different, but the quasi-shapes of these pockets are the most similar in this dendrogram and their abilities to bind seven inhibitors are very similar. Many kinases with similar sequences, for example CDK2 and CDK5 or DAPK2 and DAPK3, also have very similar quasi-shapes and inhibition profiles. This quasi-shape-based dendrogram provides a way of visualising relationships among kinases, complementing that of the classical sequence-based dendrogram. My dendrogram (Figure 33) demonstrates that the similarity in quasi-shape can sometimes explain the ability to bind a set of ligands regardless of the overall sequence identity.

### *3.3.5 A case study: BUB1 kinase*

I illustrate the use of this method with a homology model of BUB1 kinase. The information I used to construct the dendrogram is based on the information about the shape from a homology model alone. The shape of the BUB1's active site was compared with 37 non-redundant kinase structures. The method employs the centre point of 17 active site residues to construct the distance matrices for each kinase and then to find the correlation between them. The homology model of BUB1 appears to have highest similarity with CLK1. If one discards all the tyrosine kinases, it is apparent that CLK1, DAPK2, DAPK3, and PIM1 appear to be in the same group in the quasi-shape classification.

There is evidence from a study by the Structural Genomics Consortium (Marsden & Knapp, 2008) that a class of imidazo-pyridazine inhibitors which binds to CLK1, PIM1 and DAPK3 shows selectivity against a panel of 40 Ser/Thr kinases. PIM1 and CLK1 share only 18% sequence identity but they appear in the same branch of the shape-based dendrogram if tyrosine kinases are not considered. Therefore, my quasi-shape analysis is useful for selecting kinases which share similarity in the shape of the adenine binding pocket. Since BUB1 appears in the shape based dendrogram to be in the same branch as CLK1, PIM1, DAPK2, and DAPK3, an inhibitor of BUB1 might be designed by using a CLK1 inhibitor with alteration in electrostatic properties to suit BUB1.

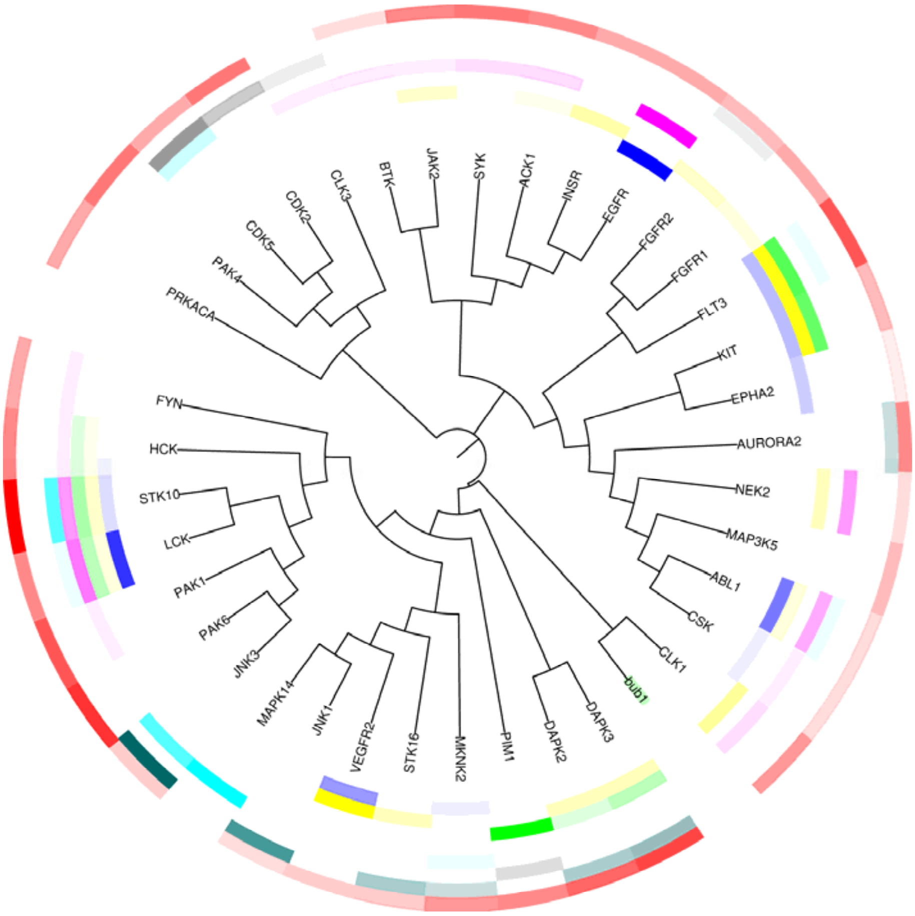


Figure 34. BUB1 and its similarity to other kinase with available inhibition constants.

### 3.4 Conclusion

For the kinase pairs which show highest similarity in their quasi-shape, STK10 and LCK, there is a correlation between the similarity in the quasi-shape of the pocket and the ligand selectivity. An experimental analysis of CLK1, PIM1 and DAPK3 suggests that my novel shape comparison method is able to distinguish kinases which bind to the same compound. The method may be useful for predicting the binding characteristics of kinases of unknown structure, as illustrated in this chapter for the BUB1 kinase domain. However, in order to understand the determinants of inhibitor selectivity, the electrostatic properties of each amino acid cannot be neglected. In the following chapter, I investigate this further in order to answer the question: Which residues in the active site influence how well the kinase can bind staurosporine?

# Chapter 4

## 4 UNDERSTANDING INFLUENTIAL RESIDUES USING QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIPS

A reverse pharmacophore approach.

*Amino acid residues in the active site that influence how well staurosporine can bind are investigated by a Multiple Linear Regression method. This approach resembles QSAR, the difference being that, instead of correlating the differences in functional groups of the ligand with the $K_d$ of the same protein, the differences in the distances between all the side chains of different kinases are chosen in order to identify those that correlate with the $K_d$ for staurosporine. A set of "influential" residues, where differences in spatial arrangement have been shown to affect the $K_d$ tremendously, are obtained from the set of distances that have the highest correlation with the $K_d$. A phylogenetic tree built from these 13 influential residues alone can cluster the kinase inhibition profiles as successfully as the general sequence alignment phylogenetic tree. The QSAR equation can be interpreted as a preference for shorter or longer distances between these influential residues, and is shown to agree well with the $K_d$. Closure between the N-terminal and C-terminal lobe and a larger size of the gatekeeper residue promote tight binding to staurosporine.*

### 4.1 Introduction

Staurosporine has very few strongly electrostatic features. Its major interactions with protein kinases are largely steric with non-polar groups. Thus, I hypothesised that the tightness of inhibitor binding might be determined by the compactness of residues in the pocket. In order to test this idea and to predict binding affinities from the structures, I assumed that good binding requires certain geometric restraints and investigated which distance descriptors correlate well with the dissociation constants. For instance, if the distances are shorter for most of the structures with low binding constants, this would suggest that contraction along that direction is required for tight binding. The goal for this experiment is to predict binding affinities directly from structures. This approach resembles a Quantitative Structure Activity Relationship (QSAR), but all the input parameters are measured from the structure in terms of distances that constitute the quasi-shape of the ATP binding pocket.

Although QSAR methodologies have been widely used in order to try to understand binding affinities through various parameters related to lipophilicity, charges and hydrogen bonding character, distances between certain atoms in the protein have not been used. Because the experimental data depend very much on the method and the experimentalist, I chose dissociation constants of staurosporine ($K_{d,STU}$) from Fabian *et al.* (Fabian et al, 2005) as the sole source of my experimental binding data. Structures in this training set have $K_{d,STU}$ between 0.5 to 870 nM and both adenine-containing or staurosporine-bound structures are considered. Adenine ring-containing structures are included in the data set on the assumption that the rigid parts of the pockets that harbour adenosine or staurosporine share similar conformations and electronic features. The advantage of assuming that the structure of the adenosine phosphate-bound complex resembles the same enzyme in the staurosporine-bound complex is that there are more structures in complex with adenosine containing compounds. The greater number of structures with available $K_{d,STU}$ values allowed me to test my equation by predicting $K_{d,STU}$ for further kinase structures co-crystallised with adenine ring containing ligands.

## 4.2 Methods

### *4.2.1 Dataset*

I filtered the Protein Data Bank (Berman et al, 2000) for X-ray structures of every kinase that Fabian *et al.* (Fabian et al, 2005) report a $K_d$ for staurosporine ($K_{d,STU}$), and selected only structures which are co-crystallised as either staurosporine or adenosine phosphate complexes.

### *4.2.2 Distance selection*

As in the case of shape analysis, distances from 15 representative atoms between all non-gapped residues surrounding the pocket were measured and written out in the form of tab-delimited file. The calculated distances are coloured in a Microsoft Excel spreadsheet based on their values as follows.

Minimum < Lower Tail < Mean – S.D.

Maximum > Upper Tail > Mean + S.D.

### *4.2.3 Multiple linear regression*

Multiple linear regression was performed using program XLSTAT (Fahmy, 2008) to find the best equation to relate the distances measured between the centre points near the end of the side chains (see Figure 11) and $\log_{10}K_{d,STU}$.

### *4.2.4 Descriptor representation*

The distances shown to correlate with the binding affinities were drawn manually in the structure of PKA (PDB ID 1stc, chain E) by writing dummy atoms and specifying the connections between them. The file was visualised using the Pymol molecular visualisation program (DeLano, 2002).

### *4.2.5 Circular dendrogram construction*

The relationships between 113 kinases in Fabian's data set were drawn based on selections of some residues in the pocket using neighbour-joining clustering in ClustalX (Thompson et al, 1997). The dendrograms were produced with gradient colour representation using program iTOL (Letunic & Bork, 2007) in order to reflect the $\log_{10}K_d$ values of the inhibitors, including:

*4.2.5.1 Clustering based on similarity of amino acid residues in contact with staurosporine and showing correlation (<-0.4 and > 0.4) with $K_{d,STU}$ in the multiple linear regression analysis. The gatekeeper residues are shown on the outer circle of the dendrogram.*

*4.2.5.2 Clustering based on similarity in the whole catalytic domain sequence. The tree was illustrated with data on the ability to bind inhibitors together with pictures of the catalytic domain of kinases from various families.*

*4.2.5.3 Clustering based on equivalent amino acid residues which are within 3 Å distance from SB202190 in the crystal structure PDB ID 1PME with the binding affinities for SB202190 ($-\log_{10}K_{d,SB202190}$)*

*4.2.5.4 Clustering based on equivalent amino acid residues which are within 3 Å distance from SP600125 in the crystal structure PDB ID 1UKI with the binding affinities for SP600125 ($-\log_{10}K_{d,SP600125}$)*

*4.2.5.5 Clustering based on equivalent amino acid residues which are within 3 Å distance from Iressa in the crystal structure PDB ID 2ITY with the binding affinities for Iressa ($-\log_{10}K_{d,Iressa}$)*

*4.2.5.6 Clustering based on equivalent amino acid residues which are within 3 Å distance from LY333531 in the crystal structure PDB*

*ID 1UU3 and 2J2I with the binding affinities for LY333531 (-$log_{10}K_{d,LY333531}$)*

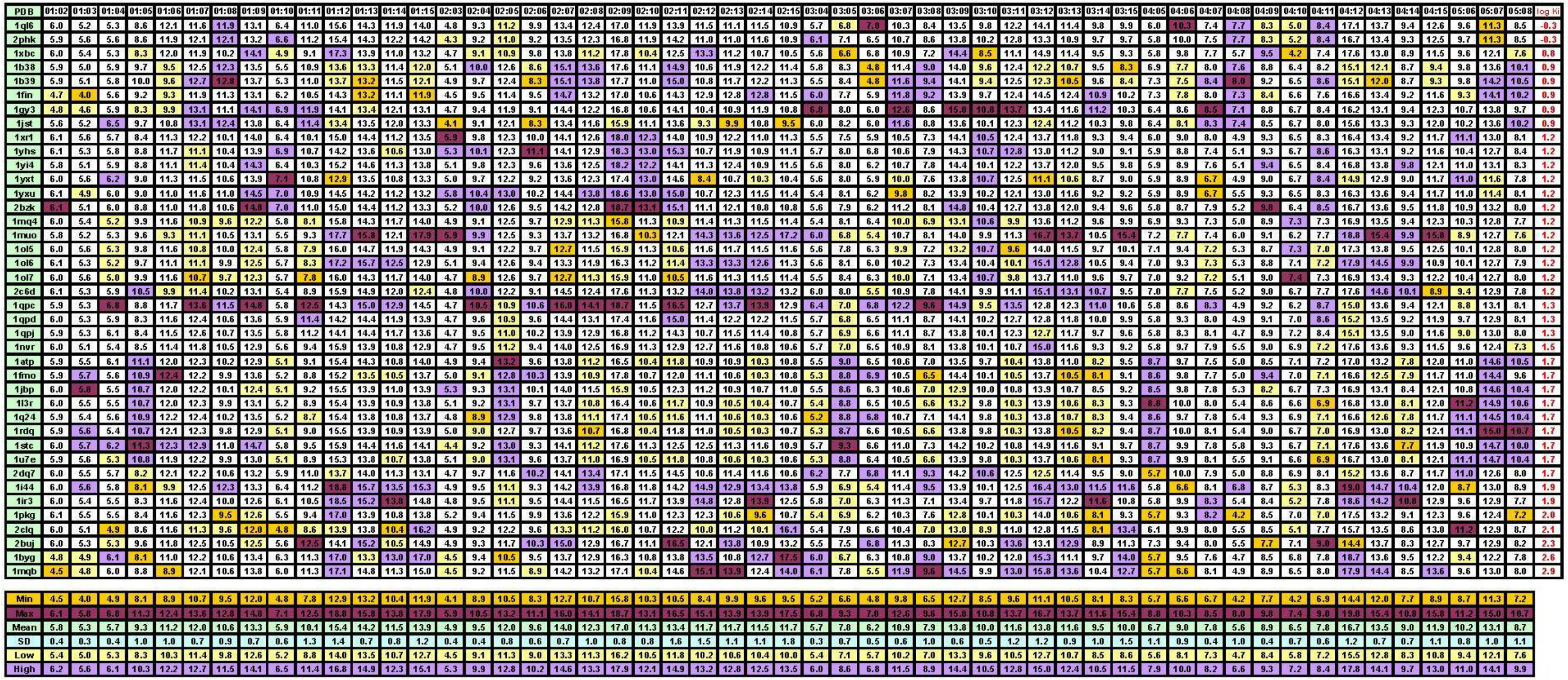
For the first two dendrograms, the selected inhibitors (i.e. staurosporine, LY-333531, SU11248, and ZD-6474) are amongst the most promiscuous ligand in the Fabian's dataset. Therefore, the number of the kinases they can bind is sufficient to demonstrate trends in binding affinities.

### 4.3 Results & Discussion

I avoided including distances that are influenced by the type of ligand bound in the structure. This was achieved by discarding residue points that differ in position when found in contact with ATP or staurosporine. In this way I could be sure that the differences in distances were independent of whether staurosporine or ATP complexes are compared. The distances between pairs of 15 frequently occurring residues are chosen for Multiple Linear Regression. These distances are measured from the centres of the distinctive parts of residues that are equivalent to these residues in cAMP dependent protein kinase (PKA).

Point 01: LEU 49
Point 02: GLY 50
Point 03: VAL 57
Point 04: ALA70
Point 05: MET 71
Point 06: LYS 72
Point 07: VAL 104
Point 08: MET 120
Point 09: GLU 121
Point 10: TYR 122
Point 11: VAL 123
Point 12: GLU 170
Point 13: ASN 171
Point 14: THR 183
Point 15: ASP 184

Table 5. Distances between pairs of 15 frequently occurring residues chosen for Multiple Linear Regression.



| PDB | 01:02 | 01:03 | 01:04 | 01:05 | 01:06 | 01:07 | 01:08 | 01:09 | 01:10 | 01:11 | 01:12 | 01:13 | 01:14 | 01:15 | 02:03 | 02:04 | 02:05 | 02:06 | 02:07 | 02:08 | 02:09 | 02:10 | 02:11 | 02:12 | 02:13 | 02:14 | 02:15 | 03:04 | 03:05 | 03:06 | 03:07 | 03:08 | 03:09 | 03:10 | 03:11 | 03:12 | 03:13 | 03:14 | 03:15 | 04:05 | 04:06 | 04:07 | 04:08 | 04:09 | 04:10 | 04:11 | 04:12 | 04:13 | 04:14 | 04:15 | 05:06 | 05:07 | 05:08 | log Ki |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1ql6 | 6.0 | 5.2 | 5.3 | 8.6 | 12.1 | 11.6 | 11.9 | 13.1 | 6.4 | 11.0 | 15.6 | 14.3 | 11.9 | 14.0 | 4.8 | 9.3 | 11.2 | 9.9 | 13.4 | 12.4 | 17.0 | 11.9 | 13.9 | 11.5 | 11.1 | 11.5 | 10.9 | 5.7 | 6.8 | 7.0 | 10.3 | 8.4 | 13.5 | 9.8 | 12.2 | 14.1 | 11.4 | 9.8 | 9.9 | 6.0 | 10.3 | 7.4 | 7.7 | 8.3 | 5.0 | 8.4 | 17.1 | 13.7 | 9.4 | 12.6 | 9.6 | 11.3 | 8.5 | -0.3 |
| 2phk | 5.9 | 5.6 | 5.6 | 8.6 | 11.9 | 12.1 | 12.1 | 13.2 | 6.6 | 11.2 | 15.4 | 14.3 | 12.2 | 14.2 | 4.3 | 9.2 | 11.0 | 9.2 | 13.5 | 12.3 | 16.8 | 11.9 | 14.2 | 11.0 | 11.0 | 11.6 | 10.9 | 6.1 | 7.1 | 6.5 | 10.7 | 8.6 | 13.8 | 10.2 | 12.8 | 13.3 | 10.9 | 9.7 | 9.7 | 5.8 | 10.0 | 7.5 | 7.7 | 8.3 | 5.2 | 8.4 | 16.7 | 13.4 | 9.3 | 12.5 | 9.7 | 11.3 | 8.5 | -0.3 |
| 1xbc | 6.0 | 5.4 | 5.3 | 8.3 | 12.0 | 11.9 | 10.2 | 14.1 | 4.9 | 9.1 | 17.3 | 13.9 | 11.0 | 13.2 | 4.7 | 9.1 | 10.9 | 9.8 | 13.8 | 11.2 | 17.8 | 10.4 | 12.5 | 13.3 | 11.2 | 10.7 | 10.5 | 5.6 | 6.6 | 6.8 | 10.9 | 7.2 | 14.4 | 8.5 | 11.1 | 14.9 | 11.4 | 9.5 | 9.3 | 5.8 | 9.8 | 7.7 | 5.7 | 9.5 | 4.2 | 7.4 | 17.6 | 13.0 | 8.9 | 11.5 | 9.6 | 12.1 | 7.6 | 0.8 |
| 1b38 | 5.9 | 5.0 | 5.9 | 9.7 | 9.5 | 12.5 | 12.3 | 13.5 | 5.5 | 10.9 | 13.6 | 13.3 | 11.4 | 12.0 | 5.1 | 10.0 | 12.6 | 8.6 | 15.1 | 13.6 | 17.6 | 11.1 | 14.9 | 10.6 | 11.9 | 12.2 | 11.4 | 5.8 | 8.3 | 4.8 | 11.4 | 9.0 | 14.0 | 9.6 | 12.4 | 12.2 | 10.7 | 9.5 | 8.3 | 6.9 | 7.7 | 8.0 | 7.6 | 8.8 | 6.4 | 8.2 | 15.1 | 12.1 | 8.7 | 9.4 | 9.8 | 13.6 | 10.1 | 0.9 |
| 1b39 | 5.9 | 5.1 | 5.8 | 10.0 | 9.6 | 12.7 | 12.8 | 13.7 | 5.3 | 11.0 | 13.7 | 13.2 | 11.5 | 12.1 | 4.9 | 9.7 | 12.4 | 8.3 | 15.1 | 13.8 | 17.7 | 11.0 | 15.0 | 10.8 | 11.7 | 12.2 | 11.3 | 5.5 | 8.4 | 4.8 | 11.6 | 9.4 | 14.1 | 9.4 | 12.5 | 12.3 | 10.5 | 9.6 | 8.4 | 7.3 | 7.5 | 8.4 | 8.0 | 9.2 | 6.5 | 8.6 | 15.1 | 12.0 | 8.7 | 9.3 | 9.8 | 14.2 | 10.5 | 0.9 |
| 1fin | 4.7 | 4.0 | 5.6 | 9.2 | 9.3 | 11.9 | 11.3 | 13.1 | 6.2 | 10.5 | 14.3 | 13.2 | 11.1 | 11.9 | 4.5 | 9.5 | 11.4 | 9.5 | 14.7 | 13.2 | 17.0 | 10.6 | 14.3 | 12.9 | 12.8 | 12.8 | 11.5 | 6.0 | 7.7 | 5.9 | 11.8 | 9.2 | 13.9 | 9.7 | 12.4 | 14.5 | 12.4 | 10.9 | 10.2 | 7.3 | 7.8 | 8.0 | 7.3 | 8.4 | 6.6 | 7.6 | 16.6 | 13.4 | 9.2 | 11.6 | 9.3 | 14.1 | 10.2 | 0.9 |
| 1gy3 | 4.8 | 4.6 | 5.9 | 8.3 | 9.9 | 13.1 | 11.1 | 14.1 | 6.9 | 11.9 | 14.1 | 13.4 | 12.1 | 13.1 | 4.7 | 9.4 | 11.9 | 9.1 | 14.4 | 12.2 | 16.8 | 10.6 | 14.1 | 10.4 | 10.9 | 11.9 | 10.8 | 6.8 | 8.0 | 6.0 | 12.6 | 8.6 | 15.0 | 10.8 | 13.7 | 13.4 | 11.6 | 11.2 | 10.3 | 6.4 | 8.6 | 8.5 | 7.1 | 8.8 | 6.7 | 8.4 | 16.2 | 13.1 | 9.4 | 12.3 | 10.7 | 13.8 | 9.7 | 0.9 |
| 1jst | 5.6 | 5.2 | 6.5 | 9.7 | 10.8 | 13.1 | 12.4 | 13.8 | 6.4 | 11.4 | 13.4 | 13.5 | 12.0 | 13.3 | 4.1 | 9.1 | 12.1 | 8.3 | 13.4 | 11.6 | 15.9 | 11.1 | 13.6 | 9.3 | 9.9 | 10.8 | 9.5 | 6.0 | 8.2 | 6.0 | 11.6 | 8.8 | 13.6 | 10.1 | 12.3 | 12.4 | 11.2 | 10.3 | 9.8 | 6.4 | 8.1 | 8.3 | 7.4 | 8.5 | 6.7 | 8.0 | 15.6 | 13.3 | 9.3 | 12.0 | 10.2 | 13.6 | 10.2 | 0.9 |
| 1xr1 | 6.1 | 5.6 | 5.7 | 8.4 | 11.3 | 12.2 | 10.1 | 14.0 | 6.4 | 10.1 | 15.0 | 14.4 | 11.2 | 13.5 | 5.9 | 9.8 | 12.3 | 10.0 | 14.1 | 12.6 | 18.0 | 12.3 | 14.0 | 10.9 | 12.2 | 11.0 | 11.3 | 5.5 | 7.5 | 5.9 | 10.5 | 7.3 | 14.1 | 10.5 | 12.4 | 13.7 | 11.8 | 9.3 | 9.4 | 6.0 | 9.0 | 8.0 | 4.9 | 9.3 | 6.5 | 8.3 | 16.4 | 13.4 | 9.2 | 11.7 | 11.1 | 13.0 | 8.1 | 1.2 |
| 1yhs | 6.1 | 5.3 | 5.8 | 8.8 | 11.7 | 11.1 | 10.4 | 13.9 | 6.9 | 10.7 | 14.2 | 13.6 | 10.6 | 13.0 | 5.3 | 10.1 | 12.3 | 11.1 | 14.1 | 12.9 | 18.3 | 13.0 | 15.3 | 10.7 | 11.9 | 10.9 | 11.1 | 5.7 | 8.0 | 6.8 | 10.6 | 7.9 | 14.3 | 10.7 | 12.8 | 13.0 | 11.2 | 9.0 | 9.1 | 5.9 | 8.8 | 7.4 | 5.1 | 9.3 | 6.7 | 8.6 | 16.3 | 13.1 | 9.2 | 11.6 | 10.4 | 12.7 | 8.4 | 1.2 |
| 1yi4 | 5.8 | 5.1 | 5.9 | 8.8 | 11.1 | 11.4 | 10.4 | 14.3 | 6.4 | 10.3 | 15.2 | 14.6 | 11.3 | 13.8 | 5.1 | 9.8 | 12.3 | 9.6 | 13.6 | 12.5 | 18.2 | 12.2 | 14.1 | 11.3 | 12.4 | 10.9 | 11.5 | 5.6 | 8.0 | 6.2 | 10.7 | 7.8 | 14.4 | 10.1 | 12.2 | 13.7 | 12.0 | 9.5 | 9.8 | 5.9 | 8.9 | 7.6 | 5.1 | 9.4 | 6.5 | 8.4 | 16.8 | 13.8 | 9.8 | 12.1 | 11.0 | 13.1 | 8.3 | 1.2 |
| 1yxt | 6.0 | 5.6 | 6.2 | 9.0 | 11.3 | 11.5 | 10.6 | 13.9 | 7.1 | 10.8 | 12.9 | 13.5 | 10.8 | 13.3 | 5.0 | 9.7 | 12.2 | 9.2 | 13.6 | 12.3 | 17.4 | 13.0 | 14.6 | 8.4 | 10.7 | 10.3 | 10.4 | 5.6 | 8.0 | 5.9 | 10.0 | 7.6 | 13.8 | 10.7 | 12.5 | 11.1 | 10.6 | 8.7 | 9.0 | 5.9 | 8.9 | 6.7 | 4.9 | 9.0 | 6.7 | 8.4 | 14.9 | 12.9 | 9.0 | 11.7 | 11.0 | 11.6 | 7.8 | 1.2 |
| 1yxu | 6.1 | 4.9 | 6.0 | 9.0 | 11.0 | 11.6 | 11.0 | 14.5 | 7.0 | 10.9 | 14.5 | 14.2 | 11.2 | 13.2 | 5.8 | 10.4 | 13.0 | 10.2 | 14.4 | 13.8 | 18.6 | 13.0 | 15.0 | 10.7 | 12.3 | 11.5 | 11.4 | 5.4 | 8.1 | 6.2 | 9.8 | 8.2 | 13.9 | 10.2 | 12.1 | 13.0 | 11.6 | 9.2 | 9.2 | 5.9 | 8.9 | 6.7 | 5.5 | 9.3 | 6.5 | 8.3 | 16.3 | 13.6 | 9.4 | 11.7 | 11.0 | 11.4 | 8.1 | 1.2 |
| 2bzk | 6.1 | 5.1 | 6.0 | 8.8 | 11.0 | 11.8 | 10.6 | 14.8 | 7.0 | 11.0 | 15.0 | 14.4 | 11.2 | 13.3 | 5.2 | 10.0 | 12.6 | 9.5 | 14.2 | 12.8 | 18.7 | 13.1 | 15.1 | 11.1 | 12.1 | 10.8 | 10.8 | 5.6 | 7.9 | 6.2 | 11.2 | 8.1 | 14.8 | 10.4 | 12.7 | 13.8 | 12.0 | 9.4 | 9.6 | 5.8 | 8.7 | 7.9 | 5.2 | 9.8 | 6.4 | 8.5 | 16.7 | 13.6 | 9.5 | 11.8 | 10.9 | 13.2 | 8.0 | 1.2 |
| 1mq4 | 6.0 | 5.4 | 5.2 | 9.9 | 11.6 | 10.9 | 9.6 | 12.2 | 5.8 | 8.1 | 15.8 | 14.3 | 11.7 | 14.0 | 4.9 | 9.1 | 12.5 | 9.7 | 12.9 | 11.3 | 15.8 | 11.3 | 10.9 | 11.4 | 11.3 | 11.5 | 11.4 | 5.4 | 8.1 | 6.4 | 10.0 | 6.9 | 13.1 | 10.6 | 9.9 | 13.6 | 11.2 | 9.6 | 9.9 | 6.9 | 9.3 | 7.3 | 5.0 | 8.9 | 7.3 | 7.3 | 16.9 | 13.5 | 9.4 | 12.3 | 10.3 | 12.8 | 7.7 | 1.2 |
| 1muo | 5.8 | 5.2 | 5.3 | 9.6 | 9.3 | 11.1 | 10.5 | 13.1 | 5.5 | 9.3 | 17.7 | 15.8 | 12.1 | 17.9 | 5.9 | 9.9 | 12.5 | 9.3 | 13.7 | 13.2 | 16.8 | 10.3 | 12.1 | 14.3 | 13.6 | 12.5 | 17.2 | 6.0 | 6.8 | 5.4 | 10.7 | 8.1 | 14.0 | 9.9 | 11.3 | 16.7 | 13.7 | 10.5 | 15.4 | 7.2 | 7.7 | 7.4 | 6.0 | 9.1 | 6.2 | 7.7 | 18.8 | 15.4 | 9.9 | 15.8 | 8.9 | 12.7 | 7.6 | 1.2 |
| 1ol5 | 6.0 | 5.6 | 5.3 | 9.8 | 11.6 | 10.8 | 10.0 | 12.4 | 5.8 | 7.9 | 16.0 | 14.7 | 11.9 | 14.3 | 4.9 | 9.1 | 12.2 | 9.7 | 12.7 | 11.5 | 15.9 | 11.3 | 10.6 | 11.6 | 11.7 | 11.5 | 11.6 | 5.6 | 7.8 | 6.3 | 9.9 | 7.2 | 13.2 | 10.7 | 9.6 | 14.0 | 11.5 | 9.7 | 10.1 | 7.1 | 9.4 | 7.2 | 5.3 | 8.7 | 7.3 | 7.0 | 17.3 | 13.8 | 9.5 | 12.5 | 10.1 | 12.8 | 8.0 | 1.2 |
| 1ol6 | 6.1 | 5.3 | 5.2 | 9.7 | 11.1 | 11.1 | 9.9 | 12.5 | 5.7 | 8.3 | 17.2 | 15.7 | 12.5 | 12.9 | 5.1 | 9.4 | 12.6 | 9.4 | 13.3 | 11.9 | 16.3 | 11.2 | 11.4 | 13.3 | 13.3 | 12.6 | 11.3 | 5.6 | 8.1 | 6.0 | 10.2 | 7.3 | 13.4 | 10.4 | 10.1 | 15.1 | 12.8 | 10.5 | 9.4 | 7.0 | 9.3 | 7.3 | 5.3 | 8.8 | 7.1 | 7.2 | 17.9 | 14.5 | 9.9 | 10.9 | 10.1 | 12.7 | 7.9 | 1.2 |
| 1ol7 | 6.0 | 5.6 | 5.0 | 9.9 | 11.6 | 10.7 | 9.7 | 12.3 | 5.7 | 7.8 | 16.0 | 14.3 | 11.7 | 14.0 | 4.7 | 8.9 | 12.6 | 9.7 | 12.7 | 11.3 | 15.9 | 11.0 | 10.5 | 11.6 | 11.3 | 11.5 | 11.4 | 5.5 | 8.4 | 6.3 | 10.0 | 7.1 | 13.4 | 10.7 | 9.8 | 13.7 | 11.0 | 9.6 | 9.7 | 7.0 | 9.2 | 7.2 | 5.1 | 9.0 | 7.4 | 7.3 | 16.9 | 13.4 | 9.3 | 12.2 | 10.4 | 12.7 | 8.0 | 1.2 |
| 2c6d | 6.1 | 5.3 | 5.9 | 10.5 | 9.9 | 11.4 | 10.2 | 13.1 | 5.4 | 8.9 | 15.9 | 14.9 | 12.0 | 12.4 | 4.8 | 10.0 | 12.2 | 9.1 | 14.5 | 12.4 | 17.6 | 11.3 | 13.2 | 14.0 | 13.8 | 13.2 | 13.2 | 6.0 | 8.0 | 5.5 | 10.9 | 7.8 | 14.1 | 9.9 | 11.1 | 15.1 | 13.1 | 10.7 | 9.5 | 7.0 | 7.7 | 7.5 | 5.2 | 9.0 | 6.7 | 7.7 | 17.6 | 14.6 | 10.1 | 8.9 | 9.4 | 12.9 | 7.8 | 1.2 |
| 1qpc | 5.9 | 5.3 | 6.8 | 8.8 | 11.7 | 13.6 | 11.5 | 14.8 | 5.8 | 12.5 | 14.3 | 15.0 | 12.9 | 14.5 | 4.7 | 10.5 | 10.9 | 10.6 | 16.0 | 14.1 | 18.7 | 11.5 | 16.5 | 12.7 | 13.7 | 13.9 | 12.9 | 6.4 | 7.0 | 6.8 | 12.2 | 9.6 | 14.9 | 9.5 | 13.5 | 12.8 | 12.3 | 11.0 | 10.6 | 5.8 | 8.6 | 8.3 | 4.9 | 9.2 | 6.2 | 8.7 | 15.0 | 13.6 | 9.4 | 12.1 | 8.8 | 13.1 | 8.1 | 1.3 |
| 1qpd | 6.0 | 5.3 | 5.9 | 8.3 | 11.6 | 12.4 | 10.6 | 13.6 | 5.9 | 11.4 | 14.2 | 14.4 | 11.9 | 13.9 | 4.7 | 9.6 | 10.9 | 9.6 | 14.4 | 13.1 | 17.4 | 11.6 | 15.0 | 11.4 | 12.2 | 12.2 | 11.5 | 5.7 | 6.8 | 6.5 | 11.1 | 8.7 | 14.1 | 10.2 | 12.7 | 12.8 | 11.8 | 10.0 | 9.9 | 5.8 | 9.3 | 8.0 | 4.9 | 9.1 | 7.0 | 8.6 | 15.2 | 13.5 | 9.2 | 11.9 | 9.7 | 12.9 | 8.1 | 1.3 |
| 1qpj | 5.9 | 5.3 | 6.1 | 8.4 | 11.5 | 12.6 | 10.7 | 13.5 | 5.8 | 11.2 | 14.1 | 14.4 | 11.7 | 13.6 | 4.7 | 9.5 | 11.0 | 10.2 | 13.9 | 12.9 | 16.8 | 11.2 | 14.3 | 10.7 | 11.5 | 11.4 | 10.8 | 5.7 | 6.9 | 6.6 | 11.1 | 8.7 | 13.8 | 10.1 | 12.3 | 12.7 | 11.7 | 9.7 | 9.6 | 5.8 | 8.3 | 8.1 | 4.7 | 8.9 | 7.2 | 8.4 | 15.1 | 13.5 | 9.0 | 11.6 | 9.0 | 13.0 | 8.0 | 1.3 |
| 1nvr | 6.0 | 5.1 | 5.4 | 8.5 | 11.4 | 11.8 | 10.5 | 12.9 | 5.6 | 9.4 | 15.9 | 13.9 | 10.8 | 12.9 | 4.7 | 9.5 | 11.2 | 9.4 | 14.0 | 12.5 | 16.9 | 11.3 | 12.9 | 12.7 | 11.6 | 10.8 | 10.6 | 5.7 | 7.0 | 6.5 | 10.9 | 8.1 | 13.8 | 10.1 | 10.7 | 15.0 | 11.6 | 9.3 | 9.2 | 5.8 | 9.7 | 7.9 | 5.5 | 9.0 | 6.9 | 7.2 | 17.6 | 13.6 | 9.3 | 11.5 | 9.6 | 12.4 | 7.3 | 1.5 |
| 1atp | 5.9 | 5.5 | 6.1 | 11.1 | 12.0 | 12.3 | 10.2 | 12.9 | 5.1 | 9.1 | 15.4 | 14.3 | 10.8 | 14.0 | 4.9 | 9.4 | 13.2 | 9.6 | 13.8 | 11.2 | 16.5 | 10.4 | 11.8 | 10.9 | 10.9 | 10.3 | 10.8 | 5.5 | 9.0 | 6.5 | 10.6 | 7.0 | 13.5 | 9.7 | 10.4 | 13.8 | 11.0 | 8.2 | 9.5 | 8.7 | 9.7 | 7.5 | 5.0 | 8.5 | 7.1 | 7.2 | 17.0 | 13.2 | 7.8 | 12.0 | 11.0 | 14.6 | 10.5 | 1.7 |
| 1fmo | 5.9 | 5.7 | 5.6 | 10.9 | 12.4 | 12.2 | 9.9 | 13.6 | 5.2 | 8.8 | 15.2 | 13.5 | 10.5 | 13.7 | 5.0 | 9.1 | 12.8 | 10.3 | 13.9 | 10.9 | 17.8 | 10.7 | 12.0 | 11.1 | 10.6 | 10.3 | 10.8 | 5.3 | 8.8 | 6.9 | 10.5 | 6.5 | 14.4 | 10.1 | 10.5 | 13.7 | 10.5 | 8.1 | 9.1 | 8.6 | 9.8 | 7.7 | 5.0 | 9.4 | 7.0 | 7.1 | 16.6 | 12.5 | 7.9 | 11.7 | 11.0 | 14.4 | 9.6 | 1.7 |
| 1jbp | 6.0 | 5.8 | 5.5 | 10.7 | 12.0 | 12.2 | 10.1 | 12.4 | 5.1 | 9.2 | 15.5 | 13.9 | 11.0 | 13.9 | 5.3 | 9.3 | 13.1 | 10.1 | 14.0 | 11.5 | 15.9 | 10.5 | 12.3 | 11.2 | 10.9 | 10.7 | 11.0 | 5.5 | 8.6 | 6.3 | 10.6 | 7.0 | 12.9 | 10.0 | 10.8 | 13.9 | 10.7 | 8.5 | 9.2 | 8.5 | 9.8 | 7.8 | 5.3 | 8.2 | 6.7 | 7.3 | 16.9 | 13.1 | 8.4 | 12.1 | 10.8 | 14.6 | 10.4 | 1.7 |
| 1l3r | 6.0 | 5.5 | 5.5 | 10.7 | 12.0 | 12.3 | 9.9 | 13.1 | 5.2 | 8.9 | 15.4 | 13.9 | 10.9 | 13.8 | 5.1 | 9.2 | 13.1 | 9.7 | 13.7 | 10.8 | 16.4 | 10.6 | 11.7 | 10.9 | 10.5 | 10.4 | 10.7 | 5.4 | 8.8 | 6.5 | 10.5 | 6.6 | 13.2 | 9.8 | 10.3 | 13.9 | 10.6 | 8.3 | 9.3 | 8.8 | 10.0 | 8.0 | 5.4 | 8.6 | 6.6 | 6.9 | 16.8 | 13.0 | 8.1 | 12.0 | 11.2 | 14.9 | 10.6 | 1.7 |
| 1q24 | 5.9 | 5.4 | 5.6 | 10.9 | 12.2 | 12.4 | 10.2 | 13.5 | 5.2 | 8.7 | 15.4 | 13.8 | 10.8 | 13.7 | 4.8 | 8.9 | 12.9 | 9.8 | 13.8 | 11.1 | 17.1 | 10.5 | 11.6 | 11.1 | 10.6 | 10.3 | 10.6 | 5.2 | 8.8 | 6.8 | 10.7 | 7.1 | 14.1 | 9.8 | 10.3 | 13.8 | 10.7 | 8.3 | 9.4 | 8.6 | 9.7 | 7.8 | 5.4 | 9.3 | 6.9 | 7.1 | 16.6 | 12.6 | 7.8 | 11.7 | 11.1 | 14.5 | 10.4 | 1.7 |
| 1rdq | 5.9 | 5.6 | 5.4 | 10.7 | 12.1 | 12.3 | 9.8 | 12.9 | 5.1 | 9.0 | 15.5 | 13.9 | 10.9 | 13.9 | 5.0 | 9.0 | 12.7 | 9.7 | 13.6 | 10.7 | 16.8 | 10.4 | 11.8 | 11.0 | 10.5 | 10.3 | 10.7 | 5.3 | 8.7 | 6.6 | 10.5 | 6.6 | 13.8 | 9.8 | 10.3 | 13.8 | 10.5 | 8.2 | 9.4 | 8.7 | 10.0 | 8.1 | 5.4 | 9.0 | 6.7 | 7.0 | 16.9 | 13.0 | 8.2 | 12.1 | 11.1 | 15.0 | 10.7 | 1.7 |
| 1stc | 6.0 | 5.7 | 6.2 | 11.3 | 12.3 | 12.9 | 11.0 | 14.7 | 5.8 | 9.5 | 15.9 | 14.4 | 11.6 | 14.1 | 4.4 | 9.2 | 13.0 | 9.3 | 14.1 | 11.2 | 17.6 | 11.3 | 12.5 | 12.5 | 11.3 | 11.6 | 10.9 | 5.7 | 9.3 | 6.6 | 11.0 | 7.3 | 14.2 | 10.2 | 10.8 | 14.9 | 11.6 | 9.1 | 9.7 | 8.7 | 9.9 | 7.8 | 5.8 | 9.3 | 6.7 | 7.1 | 17.6 | 13.6 | 7.7 | 11.9 | 10.9 | 14.7 | 10.0 | 1.7 |
| 1u7e | 5.9 | 5.6 | 5.3 | 10.8 | 11.9 | 12.2 | 9.9 | 13.0 | 5.1 | 8.9 | 15.3 | 13.8 | 10.7 | 13.8 | 5.1 | 9.0 | 13.1 | 9.6 | 13.7 | 11.0 | 16.9 | 10.5 | 11.8 | 10.8 | 10.6 | 10.3 | 10.6 | 5.3 | 8.8 | 6.4 | 10.5 | 6.6 | 13.9 | 9.8 | 10.3 | 13.7 | 10.6 | 8.1 | 9.3 | 8.7 | 9.9 | 8.1 | 5.5 | 9.1 | 6.6 | 6.9 | 16.7 | 13.0 | 8.1 | 12.1 | 11.1 | 14.7 | 10.4 | 1.7 |
| 2dq7 | 6.0 | 5.5 | 5.7 | 8.2 | 12.1 | 12.2 | 10.6 | 13.2 | 5.9 | 11.0 | 13.7 | 14.0 | 11.3 | 13.1 | 4.7 | 9.7 | 11.6 | 10.2 | 14.1 | 13.4 | 17.1 | 11.5 | 14.5 | 10.6 | 11.4 | 11.6 | 10.6 | 6.2 | 7.7 | 6.8 | 11.1 | 9.3 | 14.2 | 10.6 | 12.5 | 12.5 | 11.4 | 9.5 | 9.0 | 5.7 | 10.0 | 7.9 | 5.0 | 8.8 | 6.9 | 8.1 | 15.2 | 13.6 | 8.7 | 11.7 | 11.0 | 12.6 | 8.0 | 1.7 |
| 1i44 | 6.0 | 5.6 | 5.8 | 8.1 | 9.9 | 12.5 | 12.3 | 13.3 | 6.4 | 11.2 | 18.8 | 15.7 | 13.5 | 15.3 | 4.9 | 9.5 | 11.1 | 9.3 | 14.2 | 13.9 | 16.8 | 11.8 | 14.2 | 14.9 | 12.9 | 13.4 | 13.8 | 5.9 | 6.9 | 5.4 | 11.4 | 9.5 | 13.9 | 10.1 | 12.5 | 16.4 | 13.0 | 11.5 | 11.6 | 5.8 | 6.6 | 8.1 | 6.8 | 8.7 | 5.3 | 8.3 | 19.0 | 14.7 | 10.4 | 12.0 | 8.7 | 13.0 | 8.9 | 1.9 |
| 1ir3 | 6.0 | 5.4 | 5.5 | 8.3 | 11.6 | 12.4 | 10.0 | 12.6 | 6.1 | 10.5 | 18.5 | 15.2 | 13.8 | 14.8 | 4.8 | 9.5 | 11.1 | 9.5 | 14.4 | 11.5 | 16.5 | 11.7 | 13.9 | 14.8 | 12.8 | 13.9 | 12.5 | 5.8 | 7.0 | 6.3 | 11.3 | 7.1 | 13.4 | 9.7 | 11.8 | 15.7 | 12.2 | 11.6 | 10.8 | 5.8 | 9.9 | 8.3 | 5.4 | 8.4 | 5.2 | 7.8 | 18.6 | 14.2 | 10.8 | 12.9 | 9.6 | 12.9 | 7.7 | 1.9 |
| 1pkg | 6.1 | 5.5 | 5.5 | 8.4 | 11.6 | 12.3 | 9.5 | 12.6 | 5.5 | 9.4 | 17.0 | 13.9 | 10.8 | 13.8 | 5.2 | 9.4 | 11.5 | 9.9 | 13.6 | 12.2 | 15.9 | 11.0 | 12.3 | 12.3 | 10.6 | 9.6 | 10.7 | 5.4 | 6.9 | 6.2 | 10.3 | 7.6 | 12.8 | 10.1 | 10.3 | 14.0 | 10.6 | 8.1 | 9.3 | 5.7 | 9.3 | 8.2 | 4.2 | 8.5 | 7.0 | 7.0 | 17.5 | 13.2 | 9.1 | 12.3 | 9.6 | 12.4 | 7.2 | 2.0 |
| 2clq | 6.0 | 5.1 | 4.9 | 8.6 | 11.6 | 11.3 | 9.6 | 12.0 | 4.8 | 8.6 | 13.9 | 13.8 | 10.4 | 16.2 | 4.9 | 9.2 | 12.2 | 9.6 | 13.3 | 11.2 | 16.0 | 10.7 | 12.2 | 10.0 | 11.2 | 10.1 | 16.1 | 5.4 | 7.9 | 6.6 | 10.4 | 7.0 | 13.0 | 8.9 | 11.0 | 12.8 | 11.5 | 8.1 | 13.4 | 6.1 | 9.9 | 8.0 | 5.5 | 8.5 | 5.1 | 7.7 | 15.7 | 13.5 | 8.6 | 13.0 | 11.2 | 12.9 | 8.7 | 2.1 |
| 2buj | 6.0 | 5.3 | 5.3 | 9.6 | 11.8 | 12.5 | 10.5 | 12.5 | 5.6 | 12.5 | 14.1 | 15.2 | 10.5 | 14.9 | 4.9 | 9.3 | 11.7 | 10.3 | 15.0 | 12.9 | 16.7 | 11.1 | 16.5 | 12.1 | 13.8 | 10.9 | 13.2 | 5.5 | 7.5 | 6.8 | 11.3 | 8.3 | 12.7 | 10.3 | 13.6 | 13.1 | 12.9 | 8.9 | 11.3 | 7.3 | 9.4 | 8.0 | 5.5 | 7.7 | 7.1 | 9.0 | 14.4 | 13.7 | 8.3 | 12.7 | 9.7 | 12.9 | 8.2 | 2.3 |
| 1byg | 4.8 | 4.9 | 6.1 | 8.1 | 11.0 | 12.2 | 10.6 | 13.4 | 6.3 | 11.3 | 17.0 | 13.3 | 13.0 | 17.0 | 4.5 | 9.4 | 10.5 | 9.5 | 13.7 | 12.9 | 16.3 | 10.8 | 13.8 | 13.5 | 10.8 | 12.7 | 17.5 | 6.0 | 6.7 | 6.3 | 10.8 | 9.0 | 13.7 | 10.2 | 12.0 | 15.3 | 11.1 | 9.7 | 14.0 | 5.7 | 9.5 | 7.6 | 4.7 | 8.5 | 6.8 | 7.8 | 18.7 | 13.6 | 9.5 | 12.2 | 9.4 | 12.2 | 7.8 | 2.6 |
| 1mqb | 4.5 | 4.8 | 6.0 | 8.8 | 8.9 | 12.1 | 10.6 | 13.8 | 6.0 | 11.3 | 17.1 | 14.8 | 11.3 | 15.0 | 4.5 | 9.2 | 11.5 | 8.9 | 14.2 | 13.2 | 17.1 | 10.4 | 14.6 | 15.1 | 13.9 | 12.4 | 14.0 | 6.1 | 7.8 | 5.5 | 11.9 | 9.6 | 14.5 | 9.9 | 13.0 | 15.8 | 13.6 | 10.4 | 12.7 | 5.7 | 6.6 | 8.1 | 4.9 | 8.9 | 6.5 | 8.0 | 17.9 | 14.4 | 8.5 | 13.6 | 9.6 | 13.0 | 8.0 | 2.9 |
| Min | 4.5 | 4.0 | 4.9 | 8.1 | 8.9 | 10.7 | 9.5 | 12.0 | 4.8 | 7.8 | 12.9 | 13.2 | 10.4 | 11.9 | 4.1 | 8.9 | 10.5 | 8.3 | 12.7 | 10.7 | 15.8 | 10.3 | 10.5 | 8.4 | 9.9 | 9.6 | 9.5 | 5.2 | 6.6 | 4.8 | 9.8 | 6.5 | 12.7 | 8.5 | 9.6 | 11.1 | 10.5 | 8.1 | 8.3 | 5.7 | 6.6 | 6.7 | 4.2 | 7.7 | 4.2 | 6.9 | 14.4 | 12.0 | 7.7 | 8.9 | 8.7 | 11.3 | 7.2 | |
| Max | 6.1 | 5.8 | 6.8 | 11.3 | 12.4 | 13.6 | 12.8 | 14.8 | 7.1 | 12.5 | 18.8 | 15.8 | 13.8 | 17.9 | 5.9 | 10.5 | 13.2 | 11.1 | 16.0 | 14.1 | 18.7 | 13.1 | 16.5 | 15.1 | 13.9 | 13.9 | 17.5 | 6.8 | 9.3 | 7.0 | 12.6 | 9.6 | 15.0 | 10.8 | 13.7 | 16.7 | 13.7 | 11.6 | 15.4 | 8.8 | 10.3 | 8.5 | 8.0 | 9.8 | 7.4 | 9.0 | 19.0 | 15.4 | 10.8 | 15.8 | 11.2 | 15.0 | 10.7 | |
| Mean | 5.8 | 5.3 | 5.7 | 9.3 | 11.2 | 12.0 | 10.6 | 13.3 | 5.9 | 10.1 | 15.4 | 14.2 | 11.5 | 13.9 | 4.9 | 9.5 | 12.0 | 9.6 | 14.0 | 12.3 | 17.0 | 11.3 | 13.4 | 11.7 | 11.7 | 11.5 | 11.7 | 5.7 | 7.8 | 6.2 | 10.9 | 7.9 | 13.8 | 10.0 | 11.6 | 13.8 | 11.6 | 9.5 | 10.0 | 6.7 | 9.0 | 7.8 | 5.6 | 8.9 | 6.5 | 7.8 | 16.7 | 13.5 | 9.0 | 11.9 | 10.2 | 13.1 | 8.7 | |
| SD | 0.4 | 0.3 | 0.4 | 1.0 | 1.0 | 0.7 | 0.9 | 0.7 | 0.6 | 1.3 | 1.4 | 0.7 | 0.8 | 1.2 | 0.4 | 0.4 | 0.8 | 0.6 | 0.7 | 1.0 | 0.8 | 0.8 | 1.6 | 1.5 | 1.1 | 1.1 | 1.8 | 0.3 | 0.7 | 0.5 | 0.6 | 1.0 | 0.6 | 0.5 | 1.2 | 1.2 | 0.9 | 1.0 | 1.5 | 1.1 | 0.9 | 0.4 | 1.0 | 0.4 | 0.7 | 0.6 | 1.2 | 0.7 | 0.7 | 1.1 | 0.8 | 1.0 | 1.1 | |
| Low | 5.4 | 5.0 | 5.3 | 8.3 | 10.3 | 11.4 | 9.8 | 12.6 | 5.2 | 8.8 | 14.0 | 13.5 | 10.7 | 12.7 | 4.5 | 9.1 | 11.3 | 9.0 | 13.3 | 11.3 | 16.2 | 10.5 | 11.8 | 10.2 | 10.6 | 10.4 | 10.0 | 5.4 | 7.1 | 5.7 | 10.2 | 7.0 | 13.3 | 9.6 | 10.5 | 12.7 | 10.7 | 8.5 | 8.6 | 5.6 | 8.1 | 7.3 | 4.7 | 8.4 | 5.8 | 7.2 | 15.5 | 12.8 | 8.3 | 10.8 | 9.4 | 12.1 | 7.6 | |
| High | 6.2 | 5.6 | 6.1 | 10.3 | 12.2 | 12.7 | 11.5 | 14.1 | 6.5 | 11.4 | 16.8 | 14.9 | 12.3 | 15.1 | 5.3 | 9.9 | 12.8 | 10.2 | 14.6 | 13.3 | 17.9 | 12.1 | 14.9 | 13.2 | 12.8 | 12.5 | 13.5 | 6.0 | 8.6 | 6.8 | 11.5 | 8.9 | 14.4 | 10.5 | 12.8 | 15.0 | 12.4 | 10.5 | 11.5 | 7.9 | 10.0 | 8.2 | 6.6 | 9.3 | 7.2 | 8.4 | 17.8 | 14.1 | 9.7 | 13.0 | 11.0 | 14.1 | 9.9 | |

| PDB | 05:09 | 05:10 | 05:11 | 05:12 | 05:13 | 05:14 | 05:15 | 06:07 | 06:08 | 06:09 | 06:10 | 06:11 | 06:12 | 06:13 | 06:14 | 06:15 | 07:08 | 07:09 | 07:10 | 07:11 | 07:12 | 07:13 | 07:14 | 07:15 | 08:09 | 08:10 | 08:11 | 08:12 | 08:13 | 08:14 | 08:15 | 09:10 | 09:11 | 09:12 | 09:13 | 09:14 | 09:15 | 10:11 | 10:12 | 10:13 | 10:14 | 10:15 | 11:12 | 11:13 | 11:14 | 11:15 | 12:13 | 12:14 | 12:15 | 13:14 | 13:15 | 14:15 | (OBS) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1ql6 | 11.8 | 9.9 | 14.1 | 20.6 | 16.7 | 13.2 | 14.4 | 10.4 | 5.9 | 15.1 | 15.0 | 14.7 | 14.4 | 9.4 | 8.9 | 5.9 | 5.6 | 6.2 | 9.7 | 5.8 | 16.0 | 10.9 | 4.7 | 10.2 | 9.7 | 12.1 | 10.9 | 16.7 | 10.9 | 6.7 | 8.3 | 8.3 | 6.8 | 21.4 | 17.0 | 10.8 | 16.2 | 7.6 | 19.0 | 16.7 | 12.1 | 16.5 | 16.0 | 13.0 | 7.6 | 13.7 | 6.6 | 11.6 | 9.9 | 6.3 | 3.9 | 6.3 | -0.3 |
| 2phk | 11.8 | 10.0 | 14.0 | 20.1 | 16.4 | 13.0 | 14.3 | 10.5 | 6.1 | 15.2 | 15.0 | 15.2 | 13.4 | 8.7 | 8.5 | 5.5 | 5.7 | 6.2 | 9.9 | 6.5 | 16.1 | 10.9 | 4.7 | 10.3 | 9.8 | 12.3 | 11.5 | 16.4 | 10.7 | 6.4 | 8.2 | 8.2 | 6.5 | 21.3 | 16.8 | 10.8 | 16.3 | 7.0 | 18.8 | 16.6 | 12.3 | 16.6 | 16.8 | 13.8 | 8.7 | 14.7 | 6.7 | 11.7 | 9.8 | 6.2 | 3.8 | 6.3 | -0.3 |
| 1xbc | 13.2 | 8.8 | 13.2 | 20.9 | 16.8 | 13.5 | 14.0 | 11.0 | 6.3 | 15.4 | 13.8 | 14.5 | 13.6 | 10.2 | 9.6 | 6.2 | 5.8 | 5.8 | 9.7 | 6.4 | 15.6 | 10.3 | 5.4 | 9.5 | 9.2 | 9.7 | 9.9 | 15.9 | 11.0 | 7.4 | 8.2 | 10.2 | 8.2 | 21.3 | 16.0 | 11.0 | 15.2 | 6.3 | 19.4 | 15.1 | 10.9 | 14.6 | 16.4 | 11.9 | 7.3 | 12.7 | 5.3 | 10.3 | 8.2 | 5.2 | 4.3 | 5.7 | 0.8 |
| 1b38 | 14.0 | 11.6 | 14.8 | 20.2 | 17.1 | 14.4 | 13.2 | 10.0 | 5.8 | 13.8 | 12.9 | 13.3 | 11.7 | 8.1 | 7.7 | 4.5 | 6.4 | 5.2 | 11.5 | 6.1 | 14.3 | 9.2 | 4.6 | 7.6 | 10.1 | 13.6 | 11.4 | 14.9 | 9.4 | 6.7 | 5.1 | 10.4 | 4.8 | 18.3 | 14.0 | 9.4 | 12.4 | 7.6 | 16.2 | 15.2 | 12.0 | 14.2 | 15.1 | 12.4 | 8.1 | 12.0 | 6.5 | 9.8 | 10.2 | 4.7 | 4.6 | 4.4 | 0.9 |
| 1b39 | 14.7 | 12.0 | 15.5 | 20.4 | 17.1 | 14.6 | 13.3 | 10.2 | 6.3 | 14.0 | 12.9 | 13.4 | 11.8 | 8.0 | 7.8 | 4.6 | 6.6 | 5.3 | 11.7 | 6.2 | 14.3 | 9.2 | 4.4 | 7.6 | 10.1 | 14.0 | 11.7 | 15.4 | 9.7 | 7.0 | 5.3 | 10.8 | 5.0 | 18.4 | 14.2 | 9.3 | 12.5 | 7.8 | 16.2 | 15.2 | 12.0 | 14.2 | 15.2 | 12.4 | 8.1 | 12.1 | 6.4 | 10.0 | 10.4 | 4.9 | 4.7 | 4.6 | 0.9 |
| 1fin | 14.3 | 11.7 | 14.7 | 21.9 | 18.8 | 15.5 | 15.9 | 9.8 | 5.0 | 13.3 | 13.6 | 13.2 | 14.9 | 10.8 | 9.2 | 7.3 | 6.3 | 5.3 | 11.8 | 6.0 | 15.1 | 10.1 | 4.4 | 9.2 | 9.6 | 13.6 | 10.9 | 16.8 | 11.6 | 7.8 | 8.6 | 10.6 | 4.7 | 19.1 | 15.0 | 9.2 | 14.2 | 7.5 | 17.2 | 16.0 | 12.4 | 15.6 | 15.8 | 12.9 | 7.7 | 12.9 | 6.1 | 10.8 | 9.1 | 5.9 | 3.8 | 5.7 | 0.9 |
| 1gy3 | 13.7 | 11.2 | 14.5 | 20.8 | 18.0 | 15.0 | 16.2 | 10.0 | 4.8 | 13.9 | 14.2 | 13.7 | 13.1 | 8.9 | 8.3 | 6.1 | 6.8 | 5.3 | 11.9 | 6.4 | 15.6 | 10.5 | 4.6 | 10.0 | 10.0 | 13.3 | 11.5 | 16.2 | 11.1 | 7.6 | 8.7 | 10.7 | 5.2 | 19.5 | 15.2 | 9.3 | 15.0 | 7.6 | 16.5 | 15.5 | 12.3 | 16.1 | 16.0 | 12.9 | 8.0 | 13.8 | 6.2 | 11.2 | 9.0 | 6.0 | 3.6 | 6.2 | 0.9 |
| 1jst | 13.5 | 11.4 | 14.1 | 20.3 | 18.1 | 14.9 | 15.9 | 9.2 | 4.5 | 12.8 | 13.8 | 12.7 | 12.7 | 9.3 | 8.1 | 6.2 | 6.2 | 4.9 | 11.6 | 5.6 | 14.8 | 10.5 | 4.4 | 9.6 | 9.3 | 13.4 | 10.5 | 15.3 | 11.0 | 7.3 | 8.2 | 10.2 | 4.2 | 18.5 | 14.8 | 8.7 | 14.2 | 7.5 | 16.4 | 15.7 | 12.0 | 15.9 | 15.6 | 12.9 | 7.4 | 13.2 | 5.5 | 10.5 | 8.4 | 6.1 | 3.7 | 6.2 | 0.9 |
| 1xr1 | 13.3 | 9.0 | 13.7 | 21.0 | 18.2 | 14.6 | 15.5 | 9.5 | 7.1 | 14.7 | 15.3 | 14.8 | 13.1 | 9.2 | 8.2 | 5.2 | 5.5 | 6.4 | 12.5 | 8.0 | 14.9 | 9.4 | 5.2 | 8.9 | 8.2 | 10.9 | 10.3 | 16.9 | 12.1 | 8.3 | 9.6 | 11.1 | 7.2 | 20.5 | 15.6 | 11.1 | 15.2 | 7.7 | 19.7 | 18.0 | 13.6 | 17.3 | 16.6 | 13.8 | 9.4 | 14.6 | 6.7 | 9.7 | 9.5 | 4.8 | 4.4 | 5.6 | 1.2 |
| 1yhs | 13.2 | 8.8 | 13.7 | 20.7 | 17.9 | 14.5 | 15.3 | 10.7 | 6.2 | 14.3 | 15.4 | 15.2 | 14.7 | 10.5 | 9.2 | 6.2 | 6.0 | 5.5 | 11.1 | 5.7 | 15.1 | 10.1 | 5.8 | 10.3 | 8.2 | 11.3 | 10.2 | 16.8 | 11.9 | 8.3 | 9.4 | 10.8 | 6.0 | 20.5 | 15.6 | 11.3 | 15.4 | 8.2 | 19.8 | 17.9 | 13.7 | 17.5 | 17.2 | 14.0 | 9.7 | 15.1 | 6.6 | 9.4 | 9.6 | 4.5 | 4.7 | 5.9 | 1.2 |
| 1yi4 | 13.0 | 8.5 | 13.8 | 21.4 | 18.6 | 15.1 | 15.8 | 10.7 | 7.0 | 14.9 | 15.1 | 14.8 | 13.4 | 9.7 | 8.6 | 5.5 | 6.2 | 6.4 | 11.7 | 5.7 | 14.9 | 10.1 | 5.8 | 10.4 | 8.1 | 11.2 | 10.0 | 16.9 | 12.1 | 8.6 | 9.7 | 11.6 | 7.4 | 21.3 | 16.4 | 12.2 | 15.9 | 8.6 | 20.4 | 18.6 | 14.5 | 18.0 | 16.9 | 14.0 | 9.6 | 15.0 | 6.7 | 9.3 | 9.8 | 4.5 | 4.8 | 6.0 | 1.2 |
| 1yxt | 13.1 | 8.6 | 13.6 | 19.0 | 17.6 | 14.3 | 15.4 | 9.9 | 7.1 | 14.5 | 15.4 | 14.9 | 11.4 | 8.7 | 8.2 | 5.3 | 5.0 | 4.8 | 11.2 | 6.6 | 15.1 | 10.9 | 6.3 | 10.3 | 8.2 | 11.1 | 10.2 | 15.7 | 12.0 | 8.4 | 9.9 | 10.9 | 5.7 | 19.4 | 15.4 | 10.7 | 15.1 | 8.4 | 18.9 | 18.0 | 13.9 | 17.7 | 16.6 | 14.0 | 9.4 | 14.8 | 5.9 | 9.1 | 8.2 | 4.8 | 4.0 | 5.8 | 1.2 |
| 1yxu | 13.0 | 8.9 | 13.7 | 20.9 | 18.4 | 14.8 | 15.3 | 9.5 | 7.3 | 14.4 | 15.2 | 14.7 | 12.9 | 9.5 | 8.7 | 5.1 | 4.5 | 5.4 | 11.3 | 7.0 | 15.7 | 10.9 | 6.4 | 9.8 | 7.8 | 11.4 | 10.3 | 17.2 | 12.7 | 9.0 | 9.8 | 11.3 | 6.7 | 20.5 | 15.8 | 11.1 | 15.1 | 8.3 | 19.9 | 18.3 | 13.8 | 17.5 | 16.8 | 13.8 | 8.9 | 14.6 | 6.3 | 9.5 | 9.3 | 5.0 | 4.9 | 6.3 | 1.2 |
| 2bzk | 13.4 | 8.5 | 13.6 | 21.3 | 18.4 | 14.7 | 15.6 | 10.8 | 6.9 | 14.8 | 15.0 | 14.7 | 13.5 | 9.7 | 8.1 | 5.5 | 6.8 | 6.0 | 11.3 | 5.5 | 15.1 | 10.3 | 6.0 | 10.5 | 8.5 | 10.9 | 10.2 | 17.3 | 12.5 | 8.8 | 10.0 | 11.4 | 6.7 | 21.1 | 16.1 | 11.9 | 15.8 | 8.2 | 20.2 | 18.3 | 14.0 | 17.5 | 17.2 | 14.0 | 9.8 | 14.9 | 6.5 | 9.4 | 9.3 | 4.5 | 4.6 | 5.5 | 1.2 |
| 1mq4 | 14.3 | 12.9 | 14.2 | 21.3 | 17.8 | 14.6 | 15.3 | 10.0 | 6.4 | 14.2 | 15.8 | 12.8 | 12.8 | 8.7 | 8.3 | 5.5 | 5.6 | 4.6 | 10.9 | 5.7 | 15.1 | 10.4 | 4.4 | 9.8 | 8.7 | 11.7 | 9.2 | 16.2 | 11.5 | 7.3 | 9.3 | 10.0 | 6.2 | 19.0 | 14.7 | 8.8 | 14.4 | 6.7 | 19.0 | 17.2 | 13.1 | 17.4 | 14.5 | 11.8 | 7.2 | 12.5 | 5.9 | 11.0 | 9.1 | 6.0 | 3.9 | 6.1 | 1.2 |
| 1muo | 14.9 | 12.8 | 14.7 | 22.8 | 18.6 | 14.0 | 18.6 | 8.1 | 5.8 | 12.3 | 12.4 | 10.7 | 14.3 | 9.7 | 6.4 | 10.4 | 6.0 | 4.5 | 10.1 | 5.4 | 15.8 | 11.4 | 4.8 | 10.1 | 8.8 | 11.7 | 10.1 | 18.7 | 13.6 | 7.7 | 12.2 | 10.3 | 6.1 | 19.2 | 15.3 | 9.0 | 13.6 | 6.4 | 18.9 | 17.2 | 12.6 | 18.7 | 15.1 | 12.7 | 7.7 | 13.4 | 6.2 | 11.7 | 11.4 | 6.7 | 5.5 | 6.2 | 1.2 |
| 1ol5 | 14.5 | 13.1 | 14.0 | 21.6 | 17.8 | 14.6 | 15.3 | 10.0 | 6.5 | 14.4 | 15.9 | 12.6 | 13.6 | 9.0 | 8.4 | 5.8 | 5.5 | 4.7 | 10.9 | 5.6 | 15.6 | 10.4 | 4.4 | 9.8 | 8.7 | 12.1 | 9.1 | 16.9 | 11.7 | 7.2 | 9.4 | 10.1 | 6.5 | 19.5 | 14.8 | 8.9 | 14.5 | 6.7 | 19.1 | 17.4 | 13.3 | 17.6 | 14.6 | 11.7 | 7.2 | 12.4 | 6.9 | 11.6 | 9.9 | 6.0 | 3.8 | 6.1 | 1.2 |
| 1ol6 | 14.3 | 12.8 | 14.2 | 22.6 | 18.9 | 15.0 | 14.8 | 10.2 | 6.8 | 14.6 | 15.4 | 12.9 | 14.0 | 10.1 | 9.0 | 6.3 | 5.4 | 4.9 | 11.0 | 5.9 | 15.6 | 10.6 | 4.5 | 7.8 | 8.8 | 11.9 | 9.3 | 17.2 | 12.3 | 7.5 | 8.2 | 10.1 | 6.3 | 19.4 | 15.0 | 9.0 | 12.6 | 6.5 | 19.7 | 17.9 | 13.6 | 15.5 | 15.2 | 12.5 | 7.8 | 10.7 | 6.5 | 11.5 | 9.8 | 6.1 | 4.3 | 4.1 | 1.2 |
| 1ol7 | 14.3 | 13.1 | 14.3 | 21.7 | 17.9 | 14.6 | 15.4 | 10.0 | 6.4 | 14.5 | 15.8 | 12.8 | 13.3 | 8.7 | 8.5 | 5.5 | 5.4 | 4.9 | 11.0 | 6.1 | 15.2 | 10.4 | 4.3 | 9.7 | 8.8 | 12.0 | 9.3 | 16.4 | 11.4 | 7.0 | 9.1 | 10.2 | 6.8 | 19.2 | 15.0 | 8.9 | 14.6 | 6.5 | 18.8 | 17.1 | 13.1 | 17.4 | 14.3 | 11.8 | 7.3 | 12.6 | 6.2 | 11.3 | 9.5 | 6.2 | 3.8 | 6.2 | 1.2 |
| 2c6d | 14.5 | 12.8 | 14.7 | 22.3 | 19.0 | 15.1 | 12.0 | 8.8 | 5.4 | 13.1 | 13.0 | 11.6 | 13.8 | 10.1 | 7.7 | 5.0 | 5.6 | 4.8 | 10.5 | 5.8 | 15.2 | 11.0 | 5.1 | 5.6 | 8.8 | 11.3 | 9.5 | 17.1 | 12.8 | 8.0 | 4.8 | 10.2 | 6.3 | 19.0 | 15.3 | 9.6 | 10.2 | 6.3 | 17.9 | 16.6 | 12.7 | 13.8 | 14.5 | 12.3 | 7.7 | 10.2 | 5.6 | 10.4 | 13.8 | 5.8 | 8.7 | 4.7 | 1.2 |
| 1qpc | 14.0 | 10.2 | 14.3 | 19.0 | 17.4 | 14.1 | 14.9 | 10.4 | 7.9 | 14.5 | 14.1 | 14.4 | 13.4 | 10.6 | 9.4 | 7.1 | 5.4 | 5.5 | 11.7 | 5.7 | 13.1 | 10.2 | 4.3 | 9.5 | 7.2 | 10.3 | 8.4 | 15.8 | 13.0 | 8.0 | 10.9 | 10.8 | 4.3 | 17.5 | 15.4 | 9.4 | 14.9 | 8.7 | 16.2 | 16.4 | 12.4 | 16.3 | 14.4 | 13.2 | 7.8 | 13.7 | 4.7 | 8.9 | 8.2 | 6.0 | 4.1 | 6.2 | 1.3 |
| 1qpd | 13.9 | 11.0 | 14.3 | 19.2 | 17.4 | 14.0 | 14.7 | 10.8 | 8.8 | 15.2 | 15.6 | 15.0 | 13.0 | 10.1 | 9.4 | 6.2 | 5.3 | 5.5 | 11.9 | 5.8 | 13.3 | 10.1 | 4.3 | 9.4 | 7.2 | 10.9 | 8.6 | 16.0 | 13.0 | 8.0 | 10.8 | 10.7 | 4.4 | 17.6 | 15.2 | 9.4 | 14.9 | 8.4 | 16.7 | 16.8 | 12.8 | 16.9 | 14.5 | 13.1 | 7.9 | 13.8 | 4.8 | 9.0 | 8.5 | 5.8 | 4.4 | 6.1 | 1.3 |
| 1qpj | 13.7 | 11.0 | 14.0 | 19.2 | 17.4 | 13.8 | 14.5 | 9.5 | 7.4 | 13.8 | 15.0 | 13.7 | 13.8 | 10.5 | 8.7 | 6.4 | 5.6 | 5.7 | 12.3 | 6.0 | 13.2 | 10.1 | 4.2 | 9.1 | 7.1 | 11.0 | 8.5 | 16.3 | 13.4 | 8.1 | 10.8 | 10.7 | 4.3 | 17.6 | 15.4 | 9.3 | 14.7 | 8.4 | 16.5 | 16.8 | 12.7 | 16.6 | 14.4 | 13.2 | 7.7 | 13.5 | 4.9 | 9.1 | 8.7 | 6.1 | 4.5 | 6.0 | 1.3 |
| 1nvr | 13.6 | 11.1 | 12.9 | 21.7 | 17.4 | 13.9 | 14.0 | 11.2 | 7.8 | 15.7 | 15.7 | 13.6 | 15.8 | 10.7 | 9.7 | 6.1 | 5.9 | 5.4 | 11.8 | 5.1 | 15.9 | 10.4 | 5.6 | 9.4 | 8.8 | 12.0 | 8.9 | 18.6 | 12.9 | 8.8 | 9.6 | 10.5 | 5.0 | 19.6 | 15.1 | 10.1 | 14.6 | 7.8 | 18.3 | 16.5 | 12.4 | 16.3 | 15.2 | 11.6 | 6.7 | 12.0 | 6.8 | 10.4 | 11.2 | 5.1 | 5.0 | 5.8 | 1.5 |
| 1atp | 15.8 | 13.8 | 16.0 | 22.7 | 18.8 | 14.8 | 16.1 | 11.0 | 7.1 | 15.9 | 15.6 | 13.9 | 14.1 | 9.2 | 8.2 | 5.6 | 4.6 | 6.2 | 11.9 | 6.5 | 16.5 | 10.8 | 4.8 | 10.3 | 8.9 | 11.5 | 8.5 | 16.1 | 10.5 | 5.0 | 8.6 | 10.2 | 6.2 | 20.7 | 16.3 | 10.1 | 16.1 | 6.9 | 17.8 | 16.5 | 12.0 | 16.8 | 15.2 | 12.3 | 7.2 | 13.2 | 7.1 | 11.9 | 10.3 | 6.3 | 4.0 | 6.2 | 1.7 |
| 1fmo | 15.4 | 13.7 | 15.6 | 22.4 | 18.3 | 14.7 | 15.5 | 10.6 | 6.5 | 16.5 | 16.1 | 14.1 | 14.5 | 9.2 | 8.2 | 5.1 | 5.3 | 7.3 | 12.0 | 6.6 | 16.2 | 10.3 | 4.9 | 10.0 | 10.1 | 11.6 | 9.0 | 16.1 | 10.4 | 5.7 | 8.3 | 10.9 | 7.6 | 22.0 | 17.1 | 11.5 | 17.1 | 6.7 | 17.8 | 15.9 | 12.0 | 16.8 | 15.2 | 11.8 | 7.3 | 13.3 | 6.9 | 11.5 | 10.6 | 5.7 | 4.7 | 6.2 | 1.7 |
| 1jbp | 15.4 | 13.5 | 15.8 | 22.5 | 18.5 | 15.0 | 15.7 | 10.9 | 7.0 | 14.7 | 15.6 | 14.0 | 14.3 | 9.3 | 8.4 | 5.4 | 4.7 | 4.8 | 11.7 | 6.2 | 16.1 | 10.7 | 4.9 | 10.3 | 7.8 | 11.3 | 8.4 | 16.0 | 10.5 | 5.5 | 8.6 | 10.0 | 5.2 | 19.3 | 14.8 | 9.0 | 14.9 | 6.9 | 17.7 | 16.0 | 12.1 | 16.7 | 15.3 | 12.2 | 7.4 | 13.4 | 6.7 | 11.4 | 10.4 | 5.9 | 4.5 | 6.3 | 1.7 |
| 1l3r | 15.7 | 13.6 | 15.6 | 22.6 | 18.4 | 14.9 | 16.0 | 10.9 | 6.9 | 15.2 | 15.7 | 13.8 | 14.0 | 8.8 | 8.3 | 5.4 | 4.8 | 5.3 | 12.0 | 6.4 | 16.0 | 10.4 | 4.5 | 9.8 | 8.3 | 11.4 | 8.3 | 15.7 | 10.0 | 4.9 | 8.1 | 10.7 | 6.2 | 20.0 | 15.3 | 9.2 | 14.9 | 6.8 | 17.9 | 16.2 | 12.0 | 16.6 | 15.2 | 12.2 | 7.1 | 13.0 | 6.9 | 11.7 | 10.3 | 6.2 | 3.9 | 6.1 | 1.7 |
| 1q24 | 15.8 | 13.7 | 15.7 | 22.5 | 18.3 | 14.7 | 15.8 | 10.8 | 6.8 | 16.1 | 15.8 | 14.0 | 14.1 | 8.8 | 8.2 | 5.3 | 4.7 | 6.4 | 12.0 | 7.0 | 16.3 | 10.6 | 4.8 | 10.1 | 9.4 | 11.7 | 8.8 | 15.9 | 10.1 | 5.1 | 8.2 | 10.8 | 7.2 | 21.1 | 16.4 | 10.5 | 16.3 | 6.5 | 17.7 | 16.1 | 11.9 | 16.6 | 15.1 | 12.1 | 7.3 | 13.2 | 7.0 | 11.7 | 10.2 | 6.0 | 3.9 | 6.1 | 1.7 |
| 1rdq | 15.9 | 13.6 | 15.7 | 22.4 | 18.4 | 14.9 | 16.0 | 10.9 | 6.9 | 16.5 | 15.7 | 13.9 | 14.0 | 8.8 | 8.2 | 5.5 | 4.8 | 7.2 | 12.0 | 6.2 | 15.9 | 10.4 | 4.5 | 10.0 | 9.6 | 11.3 | 8.1 | 15.5 | 9.9 | 4.8 | 8.3 | 10.1 | 6.3 | 21.3 | 16.9 | 10.9 | 16.8 | 6.9 | 17.9 | 16.2 | 12.1 | 16.8 | 15.5 | 12.3 | 7.3 | 13.2 | 6.8 | 11.6 | 10.1 | 6.1 | 3.9 | 6.2 | 1.7 |
| 1stc | 15.1 | 13.6 | 15.8 | 24.2 | 19.9 | 14.7 | 16.6 | 11.6 | 6.7 | 15.5 | 16.0 | 14.6 | 16.3 | 11.1 | 9.4 | 6.9 | 5.7 | 5.3 | 11.7 | 6.6 | 16.9 | 11.3 | 3.4 | 9.8 | 8.9 | 12.2 | 9.7 | 17.5 | 11.7 | 5.2 | 8.3 | 11.8 | 7.9 | 21.5 | 16.5 | 8.6 | 15.0 | 6.5 | 18.2 | 16.5 | 11.9 | 16.6 | 15.4 | 12.4 | 7.0 | 12.9 | 6.9 | 13.8 | 11.5 | 8.1 | 4.9 | 6.6 | 1.7 |
| 1u7e | 15.7 | 13.6 | 15.5 | 22.4 | 18.5 | 14.7 | 16.0 | 10.9 | 6.9 | 16.5 | 15.7 | 13.8 | 13.9 | 8.9 | 8.1 | 5.5 | 4.9 | 7.1 | 12.0 | 6.3 | 15.9 | 10.4 | 4.6 | 10.1 | 9.7 | 11.5 | 8.3 | 15.7 | 10.1 | 4.9 | 8.4 | 10.2 | 6.4 | 21.1 | 16.8 | 10.8 | 16.8 | 6.8 | 17.7 | 16.2 | 12.0 | 16.7 | 15.2 | 12.2 | 7.2 | 13.1 | 6.8 | 11.6 | 10.0 | 6.1 | 3.9 | 6.2 | 1.7 |
| 2dq7 | 13.3 | 10.4 | 13.5 | 19.5 | 17.8 | 13.6 | 14.9 | 10.7 | 9.4 | 15.5 | 16.2 | 14.8 | 13.6 | 10.1 | 9.0 | 5.7 | 5.3 | 5.8 | 12.0 | 5.5 | 13.7 | 10.5 | 4.2 | 9.6 | 7.1 | 11.1 | 8.2 | 16.6 | 13.6 | 7.8 | 11.1 | 10.4 | 4.0 | 17.8 | 15.7 | 9.4 | 15.1 | 8.4 | 16.3 | 16.7 | 12.5 | 16.5 | 14.3 | 13.1 | 7.5 | 13.4 | 5.2 | 9.7 | 9.0 | 6.4 | 4.6 | 6.2 | 1.7 |
| 1i44 | 13.4 | 9.6 | 14.0 | 22.7 | 18.7 | 15.1 | 15.4 | 8.1 | 5.5 | 11.6 | 11.6 | 11.1 | 15.0 | 10.6 | 7.9 | 7.0 | 6.2 | 5.0 | 10.0 | 5.0 | 16.1 | 11.0 | 4.4 | 7.6 | 8.2 | 11.1 | 10.0 | 19.0 | 13.9 | 8.6 | 8.7 | 8.4 | 4.5 | 20.7 | 15.8 | 9.2 | 12.6 | 7.2 | 21.1 | 17.2 | 12.7 | 15.7 | 17.5 | 13.1 | 7.5 | 11.7 | 5.3 | 11.9 | 11.4 | 6.8 | 6.4 | 4.6 | 1.9 |
| 1ir3 | 13.2 | 9.7 | 13.4 | 21.9 | 17.8 | 15.1 | 15.6 | 11.9 | 7.0 | 15.5 | 14.8 | 14.8 | 14.2 | 10.8 | 10.9 | 7.8 | 6.0 | 5.1 | 10.4 | 5.3 | 15.7 | 10.5 | 4.6 | 9.8 | 8.8 | 10.0 | 9.1 | 16.4 | 11.6 | 7.7 | 9.3 | 8.2 | 3.9 | 20.2 | 15.2 | 9.5 | 14.8 | 6.9 | 21.2 | 17.0 | 13.5 | 16.7 | 17.7 | 13.1 | 8.4 | 13.4 | 5.2 | 11.4 | 8.0 | 6.2 | 3.9 | 6.1 | 1.9 |
| 1pkg | 13.1 | 11.1 | 12.7 | 20.6 | 16.6 | 13.3 | 14.6 | 9.7 | 8.3 | 14.1 | 15.5 | 12.9 | 12.6 | 9.0 | 7.9 | 5.7 | 5.5 | 5.5 | 12.6 | 5.8 | 15.1 | 9.9 | 5.7 | 9.6 | 6.8 | 10.5 | 7.3 | 17.6 | 12.7 | 8.5 | 11.3 | 10.6 | 4.5 | 19.7 | 14.7 | 10.1 | 14.8 | 7.9 | 20.4 | 16.8 | 12.8 | 17.3 | 16.6 | 12.1 | 7.5 | 12.9 | 5.2 | 9.8 | 7.6 | 4.8 | 3.9 | 5.6 | 2.0 |
| 2clq | 12.9 | 8.7 | 13.5 | 20.5 | 18.2 | 13.4 | 16.7 | 10.9 | 6.9 | 15.1 | 14.7 | 15.0 | 13.7 | 10.2 | 7.7 | 10.3 | 5.3 | 5.4 | 11.1 | 7.2 | 14.3 | 9.9 | 4.9 | 6.5 | 8.5 | 10.4 | 9.9 | 15.2 | 11.1 | 5.7 | 8.1 | 9.6 | 6.1 | 18.3 | 14.9 | 9.9 | 11.3 | 7.0 | 17.3 | 16.6 | 12.3 | 17.1 | 14.2 | 12.8 | 9.1 | 13.3 | 6.1 | 10.2 | 14.8 | 5.7 | 8.8 | 6.3 | 2.1 |
| 2buj | 12.1 | 13.0 | 15.3 | 19.9 | 18.5 | 14.0 | 16.0 | 10.3 | 7.6 | 13.5 | 15.8 | 15.1 | 13.4 | 10.6 | 8.8 | 7.1 | 4.9 | 5.2 | 12.5 | 5.8 | 13.6 | 10.4 | 6.3 | 9.8 | 6.1 | 12.0 | 9.0 | 15.4 | 12.6 | 8.1 | 10.6 | 11.0 | 4.8 | 17.6 | 15.4 | 10.4 | 14.7 | 9.9 | 15.5 | 16.9 | 11.7 | 17.5 | 15.0 | 13.6 | 8.9 | 14.3 | 5.6 | 7.5 | 8.9 | 5.6 | 4.2 | 6.3 | 2.3 |
| 1byg | 13.4 | 10.8 | 13.4 | 21.7 | 16.9 | 12.4 | 14.7 | 10.6 | 9.3 | 15.1 | 15.4 | 14.0 | 14.8 | 10.2 | 6.8 | 10.5 | 5.1 | 5.4 | 11.8 | 5.0 | 16.9 | 11.2 | 5.3 | 6.6 | 6.9 | 10.8 | 7.7 | 19.6 | 14.0 | 7.1 | 8.1 | 10.2 | 3.7 | 21.1 | 15.8 | 10.6 | 10.8 | 8.6 | 20.4 | 16.4 | 14.9 | 17.9 | 18.0 | 12.9 | 9.8 | 11.5 | 5.8 | 14.3 | 18.7 | 8.6 | 13.2 | 5.1 | 2.6 |
| 1mqb | 12.9 | 9.7 | 13.1 | 22.4 | 19.2 | 13.4 | 17.8 | 8.5 | 6.9 | 12.1 | 12.3 | 11.4 | 13.3 | 10.1 | 6.3 | 8.3 | 5.6 | 5.1 | 11.5 | 4.9 | 14.7 | 10.3 | 3.6 | 9.9 | 6.3 | 10.5 | 7.6 | 18.0 | 13.9 | 6.9 | 12.4 | 11.0 | 4.2 | 19.7 | 15.3 | 8.6 | 15.0 | 8.5 | 20.4 | 17.4 | 12.5 | 17.9 | 17.5 | 13.5 | 7.6 | 14.0 | 4.5 | 11.5 | 6.9 | 7.1 | 3.7 | 6.6 | 2.9 |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Min | 11.8 | 8.5 | 12.7 | 19.0 | 16.4 | 12.4 | 12.0 | 8.1 | 4.5 | 11.6 | 11.6 | 10.7 | 11.4 | 8.0 | 6.3 | 4.5 | 4.5 | 4.5 | 9.7 | 4.9 | 13.1 | 9.2 | 3.4 | 5.6 | 6.1 | 9.7 | 7.3 | 14.9 | 9.4 | 4.8 | 4.8 | 8.2 | 3.7 | 17.5 | 14.0 | 8.6 | 10.2 | 6.3 | 15.5 | 15.1 | 10.9 | 13.8 | 14.2 | 11.6 | 6.7 | 10.2 | 4.5 | 7.5 | 6.9 | 4.5 | 3.6 | 4.1 |
| Max | 15.9 | 13.8 | 16.0 | 24.2 | 19.9 | 15.5 | 18.6 | 11.9 | 9.4 | 16.5 | 16.2 | 15.2 | 16.3 | 11.1 | 10.9 | 10.5 | 6.8 | 7.3 | 12.6 | 8.0 | 16.9 | 11.4 | 6.4 | 10.5 | 10.1 | 14.0 | 11.7 | 19.6 | 14.0 | 9.0 | 12.4 | 11.8 | 8.2 | 22.0 | 17.1 | 12.2 | 17.1 | 9.9 | 21.2 | 18.6 | 14.9 | 18.7 | 18.0 | 14.0 | 9.8 | 15.1 | 7.1 | 14.3 | 18.7 | 8.6 | 13.2 | 6.6 |
| Mean | 13.9 | 11.3 | 14.3 | 21.3 | 18.0 | 14.4 | 15.3 | 10.2 | 6.8 | 14.6 | 14.8 | 13.8 | 13.7 | 9.6 | 8.4 | 6.3 | 5.6 | 5.5 | 11.4 | 6.0 | 15.3 | 10.4 | 4.8 | 9.4 | 8.5 | 11.6 | 9.5 | 16.6 | 11.8 | 7.2 | 9.0 | 10.4 | 5.7 | 19.8 | 15.5 | 9.9 | 14.6 | 7.5 | 18.4 | 16.7 | 12.7 | 16.6 | 15.7 | 12.8 | 8.0 | 13.2 | 6.1 | 10.7 | 10.0 | 5.9 | 4.7 | 5.9 |
| SD | 1.1 | 1.8 | 0.9 | 1.2 | 0.8 | 0.7 | 1.2 | 0.9 | 1.1 | 1.2 | 1.2 | 1.2 | 1.0 | 0.8 | 0.9 | 1.4 | 0.6 | 0.7 | 0.8 | 0.6 | 1.0 | 0.5 | 0.7 | 1.2 | 1.1 | 1.0 | 1.1 | 1.1 | 1.3 | 1.2 | 1.6 | 0.8 | 1.2 | 1.3 | 0.8 | 1.0 | 1.6 | 0.9 | 1.5 | 0.9 | 0.8 | 1.1 | 1.1 | 0.7 | 0.9 | 1.1 | 0.7 | 1.3 | 2.0 | 0.9 | 1.8 | 0.6 |
| Low | 12.8 | 9.5 | 13.4 | 20.1 | 17.2 | 13.7 | 14.2 | 9.4 | 5.7 | 13.4 | 13.7 | 12.6 | 12.7 | 8.8 | 7.6 | 4.8 | 5.0 | 4.8 | 10.6 | 5.4 | 14.3 | 9.9 | 4.1 | 8.2 | 7.5 | 10.6 | 8.3 | 15.5 | 10.5 | 5.9 | 7.4 | 9.5 | 4.5 | 18.5 | 14.8 | 9.0 | 13.0 | 6.6 | 16.8 | 15.9 | 11.8 | 15.6 | 14.6 | 12.0 | 7.1 | 12.1 | 5.4 | 9.4 | 7.9 | 5.0 | 3.0 | 5.3 |
| High | 15.0 | 13.1 | 15.2 | 22.5 | 18.8 | 15.1 | 16.5 | 11.1 | 7.8 | 15.7 | 16.0 | 14.9 | 14.7 | 10.4 | 9.3 | 7.7 | 6.2 | 6.2 | 12.2 | 6.7 | 16.2 | 10.9 | 5.5 | 10.5 | 9.6 | 12.5 | 10.6 | 17.7 | 13.0 | 8.4 | 10.6 | 11.2 | 7.0 | 21.1 | 16.3 | 10.9 | 16.2 | 8.3 | 19.9 | 17.6 | 13.5 | 17.7 | 16.8 | 13.5 | 8.8 | 14.3 | 6.8 | 12.0 | 12.0 | 6.7 | 6.5 | 6.5 |

The aim was to select distances from the quasi-shape defined by 15 points in contact with staurosporine (Table 5), where the points correlate with the binding affinities either positively or negatively. Multiple linear regression was used for this task and the equations produced are shown in Table 6. $K_{d,STU}$ is the dissociation constant of staurosporine, and the distance between residues X and Y is written as $D_{X\_Y.}$

Table 6. Equations correlating the influential distances with $\log_{10}K_{d,STU}$

| Random Test | $R^2$ training | $R^2$ test set | Equation |
|---|---|---|---|
| None | 0.6 | - | Equation 4:<br>$\log_{10}K_{d,STU}=3.4+0.1D_{50\_184}–0.4D_{120\_123}$ |
| 5 structures | 0.6 | 0.7 | Equation 5:<br>$\log_{10}K_{d,STU}=3.4+0.1D_{50\_184}–0.4D_{120\_123}$ |
| 10 structures | 0.7 | 0.7 | Equation 6:<br>$\log_{10}K_{d,STU}=3.6+0.1D_{50\_184}–0.4D_{120\_123}$ |

I tested the predictive power of these equations by leaving out randomly selected test sets. While the purpose of using multiple linear regression in this context was simply to select the set of distances that correlate well with the binding affinities, the resulting equations suggest that predictive power might be demonstrated if a larger dataset were available. All resulting equations appear to contain the same best sets of distances producing $R^2$ values for the random test sets of about 0.7 for both equations (Figures 35 & 36).

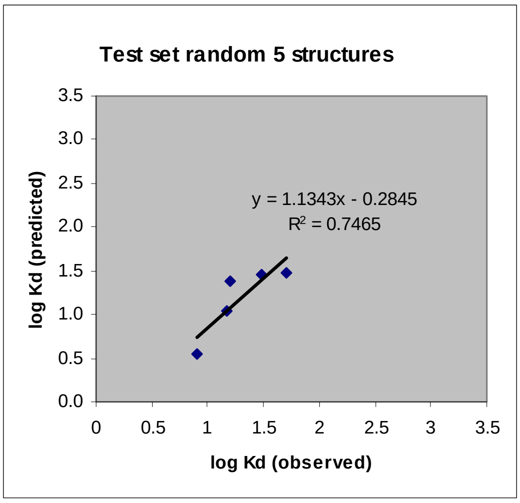


Figure 35. The predictive power of the multiple linear regression equations, tested by leaving out 5 randomly selected test sets, shows $R^2$ about 0.7.

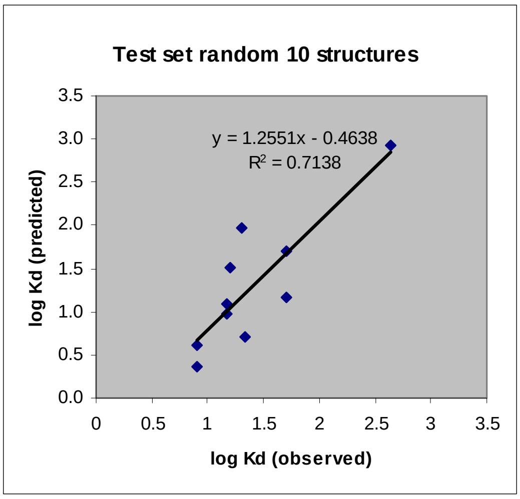


Figure 36. The predictive power of the multiple linear regression equations, tested by leaving out 10 randomly selected test sets, shows $R^2$ about 0.7.

The distance descriptors which correlate well with binding affinities, either having positive or negative influence on $K_{d,STU}$, are called the 'influential distances'. Figure 37 illustrates these influential distances in the structure of cyclic AMP dependent protein kinase (PKA), PDB ID 1stc. The residues that are used as the points of measurement for these influential distances can be used to describe how the position of their representative atoms near the end of the side chains can influence $K_{d,STU}$. From all the equations shown in Table 6, the distance between residues 50 and 184, described in the equation as $D_{50\_184}$, is directly proportional to the value of $\log_{10}K_{d,STU}$, and the distance between residues 120 and 123, $D_{120\_123}$, is inversely proportional to $\log_{10}K_{d,STU}$. The meaning of the equation is that in kinases that are tightly bound to staurosporine, i.e. have a small $\log_{10}K_{d,STU}$, there is a preference for a smaller $D_{50\_184}$ and a larger $D_{120\_123}$.

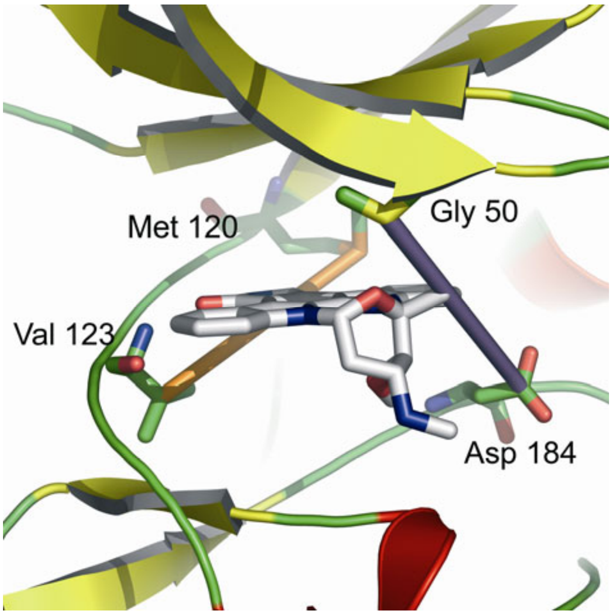


Figure 37. Interpretation of the multiple linear regression analysis shows that smaller values of $K_{d,STU}$ result from the larger size of side chains of the gatekeeper and gatekeeper+3 residues, i.e. PKA equivalent residue: $Met_{120}$ and $Val_{123}$ (orange bar). The equation suggests that the closer approach between $Gly_{50}$ of the N-terminal lobe and $Asp_{184}$ of the C-terminal lobe (purple bar) correlate with tighter binding to staurosporine.

In cAMP dependent protein kinase (PKA), the distance between residues 50 and 184 is measured between the Cα of $Gly_{50}$ of the GXGXXG motif in the N-terminal lobe to the Cγ of $Asp_{184}$ of the DFG loop in the C-terminal lobe. Staurosporine is located between the two lobes, and the closer approach of these two motifs in a direction perpendicular to the plane of staurosporine reflects the better binding affinities presumably because of the resultant tighter binding. In contrast, increasing the distance between residue 120 (gatekeeper) and 123 (gatekeeper+3) implies an expansion of the pocket along this direction. The equation suggests that these two residues should move further apart to accommodate staurosporine. The gatekeeper residue points toward the plane of staurosporine, while the gatekeeper+3 residue is located under the indolocarbazole ring. The size of the gatekeeper and the gatekeeper+3 residues may have a key role in locking the lactam in the correct orientation while making optimal steric interactions with the indolocarbazole of staurosporine. The larger size of the gatekeeper residue likely results in the larger distance and correlates with good binding because the larger volumes of the side chains in the plane of the lactam ring promote favorable hydrophobic interactions in the pocket.

Thirteen residues which are in contact with staurosporine and show correlation (<-0.4 and > 0.4) with $K_{d,STU}$ were selected for the clustering of staurosporine binding affinities. These residues are equivalent to PKA residues 49, 50, 57, 70, 71, 72, 120, 121, 122, 123, 170, 171, and 184. In Figure 38, the resulting dendrogram constructed by the neighbour-joining algorithm of the thirteen residues is combined with data on the ability to bind staurosporine from Fabian *et al.* (Fabian et al, 2005) in order to investigate whether the similarities between these influential residues would result in similar binding constant. The resulting dendrogram can cluster tight staurosporine binders into two major groups with obviously better binding affinities (Figure 38). I also display the gatekeeper residues, which in Figure 38 lie beside the affinities of staurosporine. It can be seen that the majority of staurosporine tight binders have large gatekeeper residues, e.g. Phe and Met at the position equivalent to PKA residue 120 (Figure 38). Smaller gatekeeper residues, e.g. Thr or Leu, tend to be associated with weaker binding affinities to staurosporine.

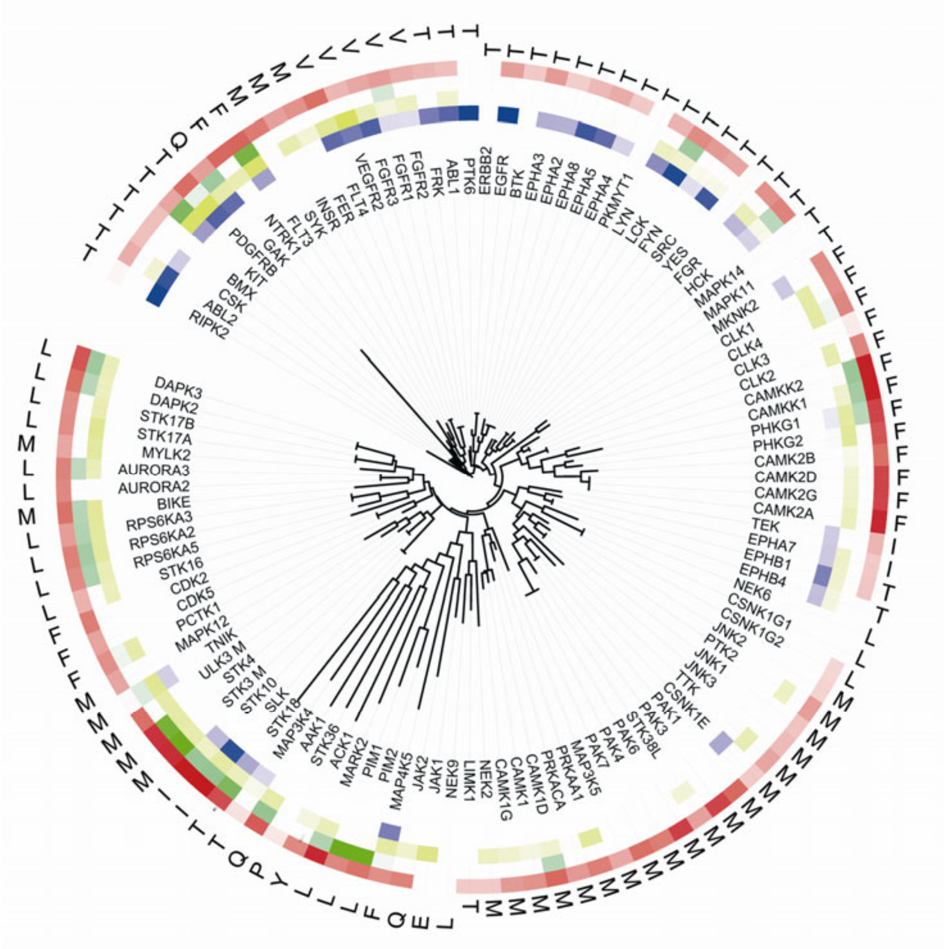


Figure 38. A dendrogram displaying relationships between 113 kinases based on neighbour joining of the 13 residues that are highly correlated to binding constants. Most kinases with better binding affinities to staurosporine (dark red) have large gatekeeper residues, e.g. phenylalanine (F), methionine (M). A majority of kinases which are inhibited by ZD-6474 (blue) has threonine (T) or valine (V) as a gatekeeper residue. Binding affinities to LY-333531 (green) and SU11248 (yellow) are shown for comparison.

The dendrogram produced from the whole catalytic domain sequence alignment was made for comparison. It is clear that the similarity in sequence of the whole catalytic kinase domain does not imply that the enzyme would bind to the same set of inhibitors (Figure 39). It is interesting that the ability to cluster ZD-6474 in the thirteen active site residue dendrogram is even better than the whole domain clustering.

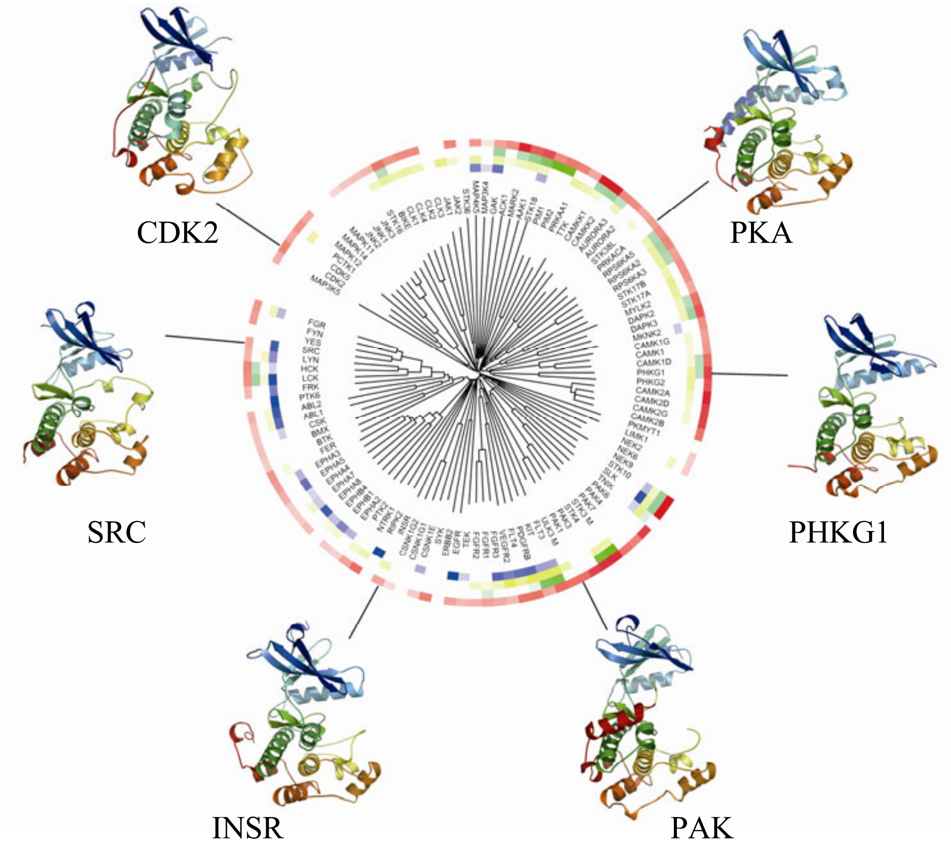


Figure 39. The binding affinities to promiscuous inhibitors, staurosporine (red), LY-333531(green), SU11248 (yellow), and ZD-6474 (blue), demonstrate that the similarity in sequence cannot predict the ability to bind inhibitor.

A further question that I address is whether there is any other inhibitor for which the active site residues can be used to distinguish good binders from poor binders. The similarities in ligand accessible residues of four promiscuous inhibitors where there are available crystal structures in complex with kinases show that we cannot predict the trend in binding affinities in this way. All of these inhibitors are more selective than staurosporine. This is probably because the inhibition by these inhibitors is governed by fewer residues. Other factors, such as the flexibility of the ligand, the electrostatic potential, and the steric interaction in the pocket might make sequence comparison of active site residues inadequate for the prediction of binding affinities.

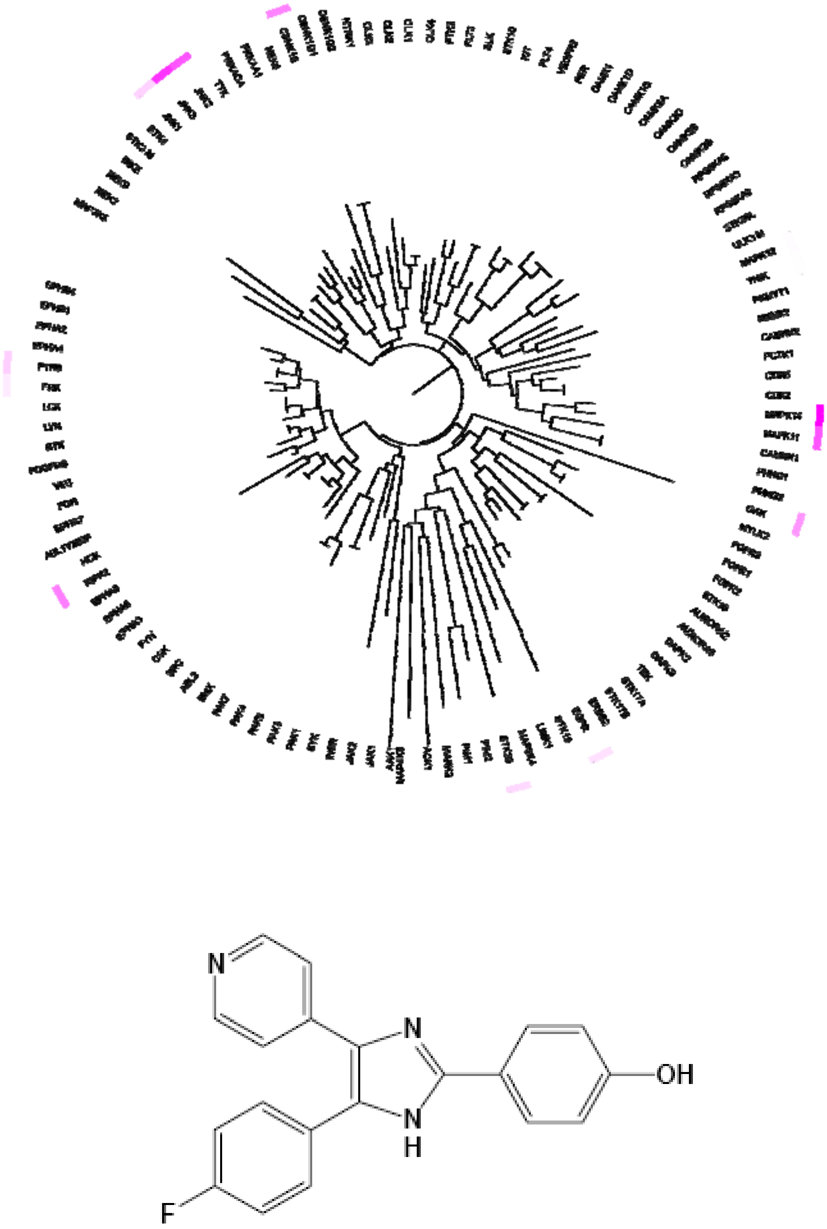


Figure 40. Dendrogram constructed from alignment of residues within SB202190 accessible region

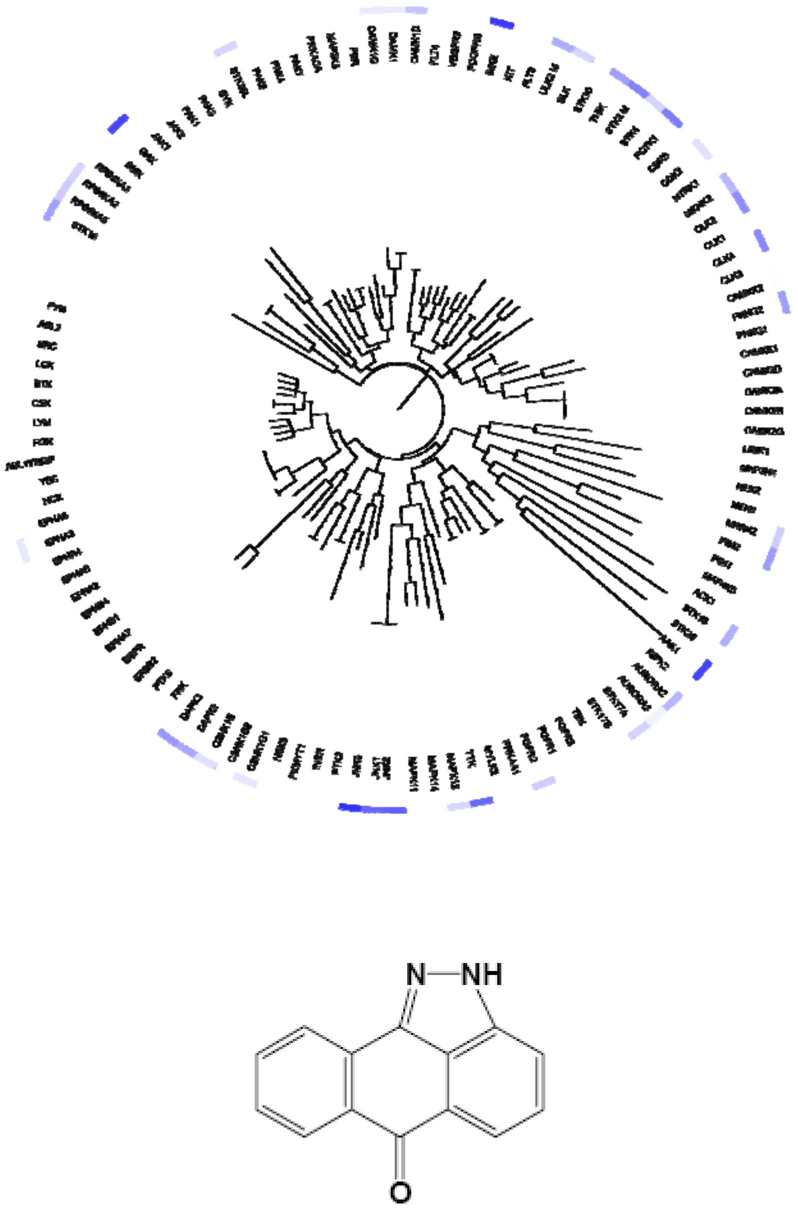


Figure 41. Dendrogram constructed from alignment of residues within SP600125 accessible region

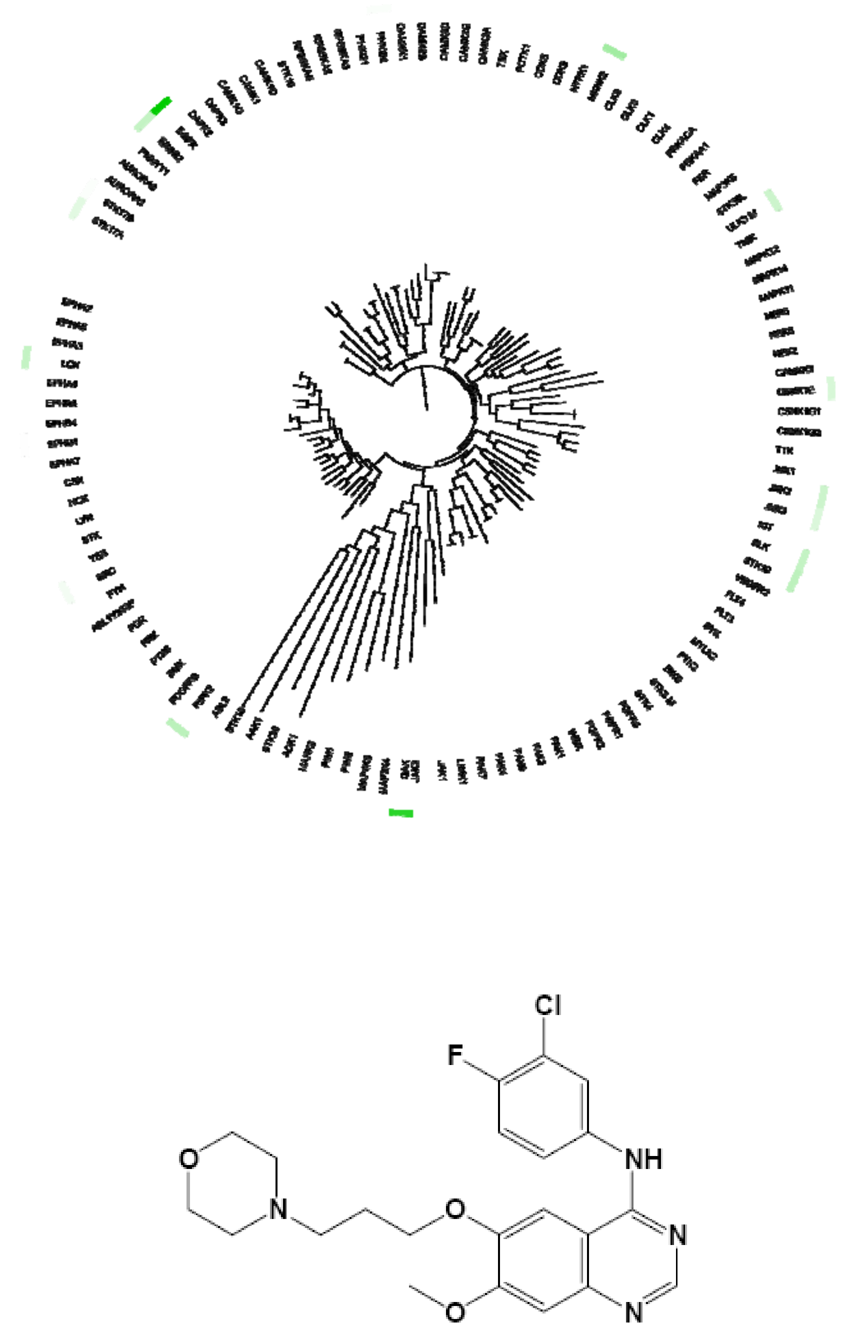


Figure 42. Dendrogram constructed from alignment of residues within Iressa accessible region

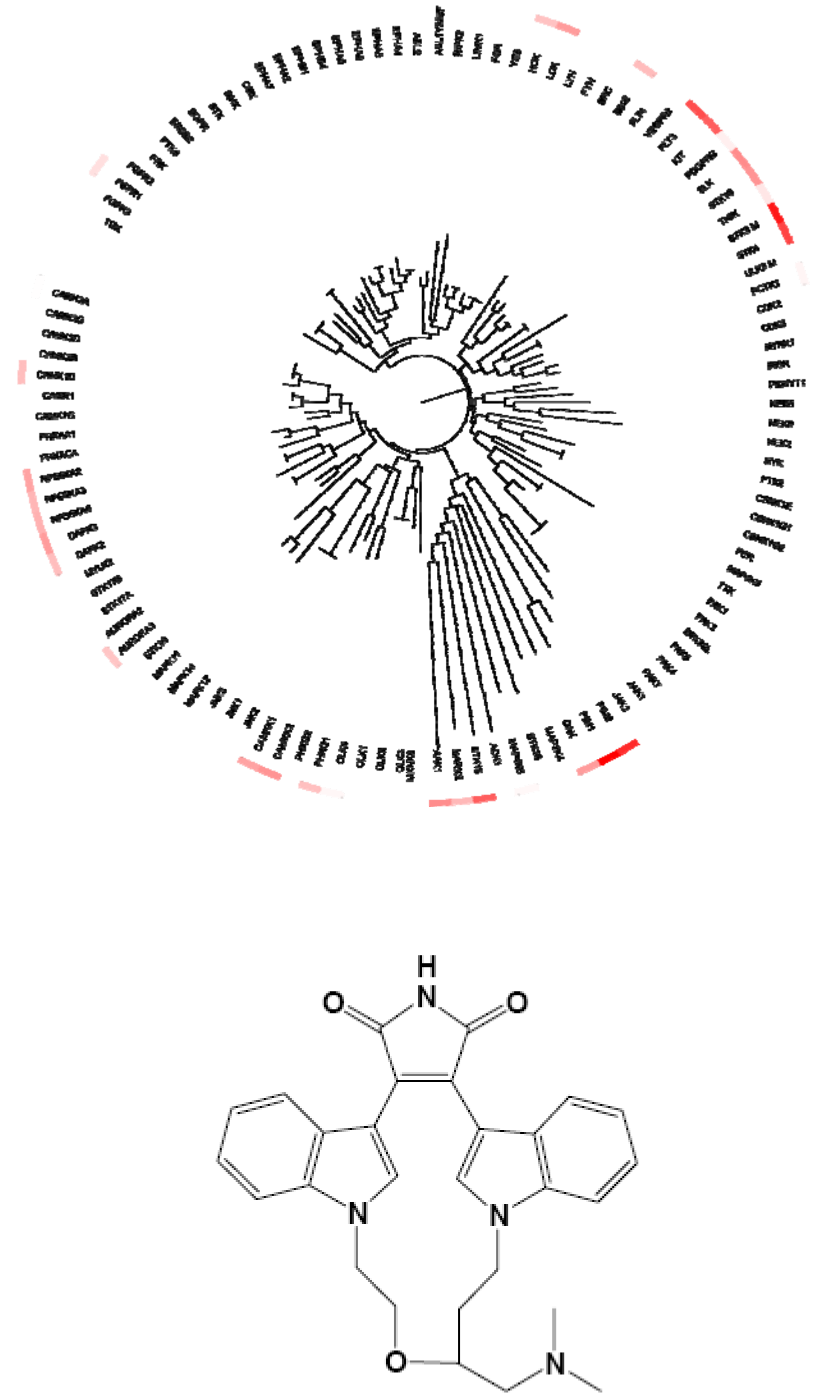


Figure 43. Dendrogram constructed from alignment of residues within LY333531 accessible region

## 4.4 Conclusion

I have dissected the contribution to kinase staurosporine binding affinities in terms of distances between residues that line the ATP binding site. I have proposed that the size of the gatekeeper and closure of the pocket together affect the tightness of staurosporine binding. Here I have shown using the three-dimensional structures of kinases that larger sizes of gatekeeper residues normally result in tighter binding to staurosporine. This is probably a result of the compactness caused by the larger volumes of the side chains in the plane of the lactam ring which result in a better fitting of staurosporine. The closure of the DFG loop and the glycine rich loop also correlates to the tightness of staurosporine binding. The clustering of kinases based on 13 influential residues also shows that the ability to bind staurosporine can be grouped roughly based on similarity of a few residues in the pocket. However, this method of clustering cannot be applied to more selective inhibitors. The amount of selectivity determining residues selected for these inhibitors for the construction of the dendrogram is not optimal, and more detailed comparison is required for each inhibitor to understand the basis of each inhibitor’s selectivity.

# Chapter 5

## 5 UNDERSTANDING CROSS-REACTIVITY BASED ON COMMON SUBSTRUCTURES

*A web application, MAHORI (http://www-cryst.bioc.cam.ac.uk/mahori), has been developed to query atomic interaction information from the protein-ligand database Credo (http://camelot.bioc.cam.ac.uk/drupal/databases/credo). When the query has been made by submitting a SMILES string, drawing the chemical functional group or typing the compound's name, the website can then display all the crucial interactions that a molecule makes with proteins in the PDB. A simple search query for staurosporine showed that the greater the number of the hydrogen bond and ionic interactions made by the methyl amine moiety, the better the binding affinity to staurosporine. This information can be useful in identifying selectivity-determining residues.*

### 5.1 Introduction

Crystal structures serve as templates for many facets of drug discovery (Breitenlechner et al, 2005). A survey of the market value of small-molecule drugs has shown that two-thirds of the sales resulted from analogue designs (Wermuth, 2006). Understanding the interactions made by chemical analogues presented in the Protein Data Bank may therefore suggest synthetic strategies for lead optimisation. Nevertheless, gathering the binding characteristics of a particular analogue is time-consuming and there is no publicly available resource that facilitates this type of understanding. My web-based application, MAHORI, is acronymed from its function for Mapping Analogous Hetero-atoms onto Residue Interactions. It aims to provide visualisation and classification of molecular interactions made by the user-query atoms obtained from the heterogen section of the PDB file (wwPDB, 2007).

A few web resources allow PDB structures and their superpositions to be queried based on ligand structure, including Relibase (Bergner et al, 2001) and IsoStar (Bruno et al, 1997). Selection of amino acid residues that interact with a queried ligand can also be achieved by FireDB (Lopez et al, 2007) and classification of ligand-protein interactions based on residue contacts can be obtained from MSDsite (Golovin et al, 2005). However, these websites do not allow for comparison of the molecular interactions of multiple structures at the level of ligand substructure.

The rationale for allowing user-defined substructure comparisons is based on the idea of the bioisostere that often shares a similar number and position of interactions with the protein, such as hydrogen bonding environment. Selection of the analogous part of the molecule could give an indication of the significant interactions that the substructure can contribute to the binding. This chapter describes an approach to selecting either the atoms that comprise the equivalent substructure in several molecules which can relate to cross-reactivity amongst several kinases or non-carbon atoms that are likely to make significant interactions which are selective for a small set of kinases.

## 5.2 Methods

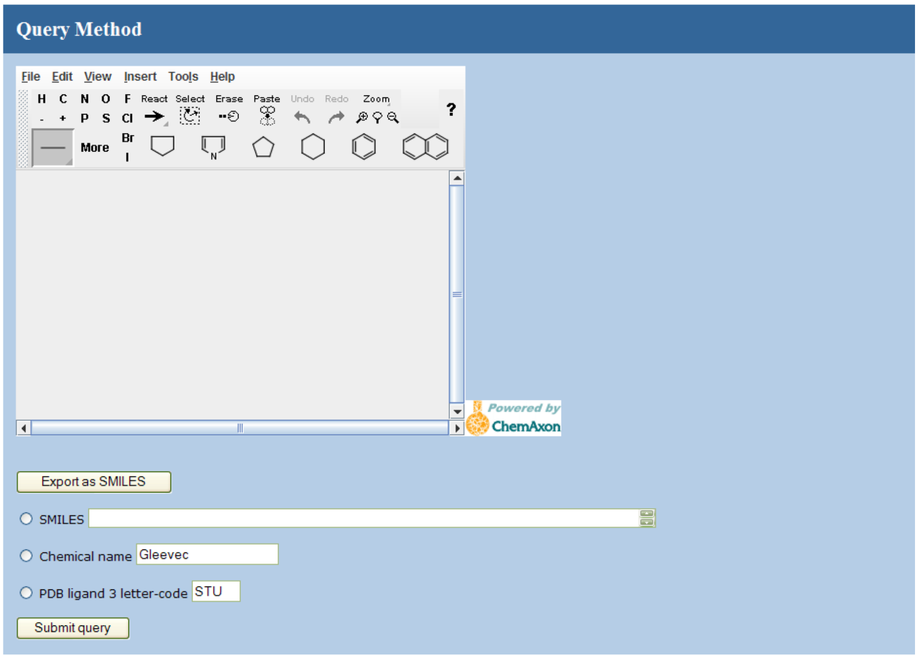
### *5.2.1 Rationale*

MAHORI supports robust querying of the Protein Data Bank through its underlying protein-ligand database, Credo (Schreyer & Blundell, 2009). This database stores all types of interactions and contact distances that the ligands make with the protein. This is achieved using a method adapted from assignment of interaction types described in the approach for optimising fragment and scaffold docking (Marcou & Rognan, 2007).

Every atom of the ligand and its contacting neighbour atoms have their pre-defined types and the distances are calculated for every interacting pair. By prioritising the types of atoms and the distances, the interaction types can be assigned for all atom pairs.

### 5.2.2 *Available query types*

The user can make a query by providing a chemical structure, a SMILES string (Weininger, 2002), a chemical name or a PDB three-letter code. The ligand's PDB three-letter code is the main type of query which will lead to a substructure selection panel. When searching by the chemical name, the program will look for the ligand's PDB three letter codes in the Credo database. When searching for a fragment, the program receives the query in the form of a chemical drawing from program MarvinSketch 4.1.12 (ChemAxonLtd., 2007) and converts it to a SMILES string.


Query Method
File Edit View Insert Tools Help
H C N O F React Select Erase Paste Undo Redo Zoom
- + P S Cl
More Br I
?
Powered by ChemAxon
Export as SMILES
Submit query

○ SMILES

○ Chemical name Gleevec

○ PDB ligand 3 letter-code STU

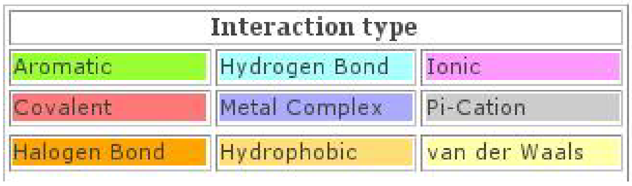
Figure 44. The MAHORI web-interface allows for various forms of ligand query, e.g. by providing a chemical structure, a SMILES string, a chemical name or a PDB three-letter code.

Once the SMILES string is obtained, the program obgrep from the OPENBABEL package (Banck et al, 2007) searches for the ligand which contains that string in the Protein Data Bank. This program will then write an output file containing a list of the ligand PDB three-letter codes.

### *5.2.3 Query execution process*

For each ligand PDB three-letter code, the program will provide the user with a list of the atoms that constitute the ligand and the name of the proteins that interact with these ligands. The user can select atoms of interest from this list. The name of the protein will be provided based on the interacting partners of those atoms. The process goes on until the last ligand in the list.

The program uses the set of ligand atoms obtained from the user to make a MySQL query for a list of interacting protein atoms. The contact distances and interaction type are provided by the Credo database. The program prints all interactions that this set of atoms make and populates the closest contact per residue into a table of residue interactions.

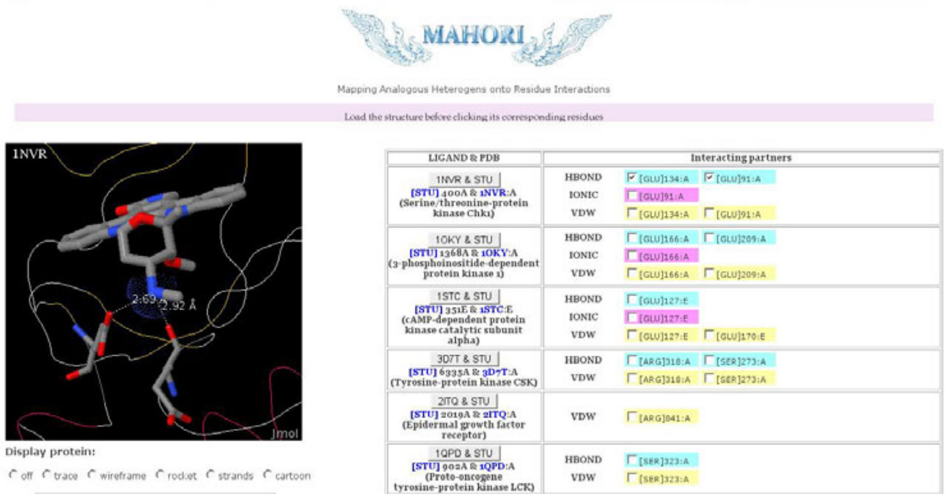


| Interaction type | | |
|---|---|---|
| Aromatic | Hydrogen Bond | Ionic |
| Covalent | Metal Complex | Pi-Cation |
| Halogen Bond | Hydrophobic | van der Waals |

Figure 45. Colour code used for displaying ligand-residue interactions in MAHORI

### *5.2.4 Data presentation*

MAHORI returns the output interactions into two panels. The left panel is a molecular viewer which can display the molecular interaction using the program Jmol 11.3.42 (TheJmolDevelopmentTeam, 2007), where the user can view the molecule in several styles, e.g. wireframe, strands, cartoon, rocket. The right panel prints out the final table for each PDB file which contains the interactions that are sorted amongst the protein atoms nearest to the ligand atoms. Interacting residues are displayed in colours according to the interaction type. Clicking the interacting residues in the right panel will trigger the picture of the interacting residues in the left panel. Multiple user-defined substructures can be retrieved and displayed in the same page.



Mapping Analogous Heterogens onto Residue Interactions

Load the structure before clicking its corresponding residues

Display protein:

○ off ○ trace ○ wireframe ○ rocket ○ strands ○ cartoon

| LIGAND & PDB | Interacting partners | | |
|---|---|---|---|
| 1NVR & STU<br>[STU] 400A & 1NVR:A<br>(Serine/threonine-protein kinase Chk1) | HBOND | ☑ [GLU]134:A | ☑ [GLU]91:A |
| | IONIC | □ [GLU]91:A | |
| | VDW | □ [GLU]134:A | □ [GLU]91:A |
| 1OKY & STU<br>[STU] 1368A & 1OKY:A<br>(3-phosphoinositide-dependent protein kinase 1) | HBOND | □ [GLU]166:A | □ [GLU]209:A |
| | IONIC | □ [GLU]166:A | |
| | VDW | □ [GLU]166:A | □ [GLU]209:A |
| 1STC & STU<br>[STU] 351E & 1STC:E<br>(cAMP-dependent protein kinase catalytic subunit alpha) | HBOND | □ [GLU]127:E | |
| | IONIC | □ [GLU]127:E | |
| | VDW | □ [GLU]127:E | □ [GLU]170:E |
| 3D7T & STU<br>[STU] 6335A & 3D7T:A<br>(Tyrosine-protein kinase CSK) | HBOND | □ [ARG]318:A | □ [SER]273:A |
| | VDW | □ [ARG]318:A | □ [SER]273:A |
| 2ITQ & STU<br>[STU] 2019A & 2ITQ:A<br>(Epidermal growth factor receptor) | VDW | □ [ARG]841:A | |
| 1QPD & STU<br>[STU] 902A & 1QPD:A<br>(Proto-oncogene tyrosine-protein kinase LCK) | HBOND | □ [SER]323:A | |
| | VDW | □ [SER]323:A | |

Figure 46. Example of MAHORI output displaying interaction made with N4′ of staurosporine. More details of N4′ interaction can be found in Figure 51.

## 5.3 Results & Discussion

### 5.3.1 *Understanding staurosporine promiscuity*

The interactions found in common between all protein kinases and staurosporine can be dissected into those that are conserved involving the lactam and tetrahydropyran moieties and the steric/hydrophobic contacts made by the indolocarbazole ring.

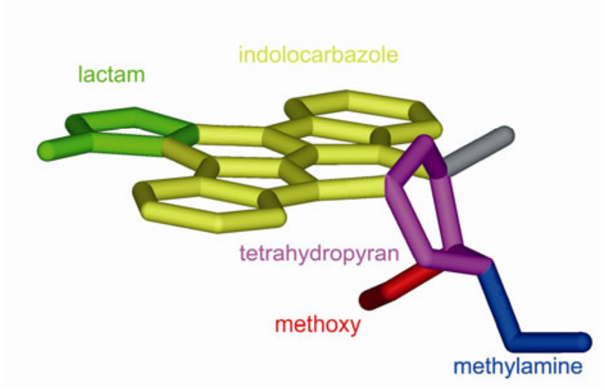


Figure 47. The potential group which likely causes alteration in binding affinities is the methylamine moiety of staurosporine.

I speculated that interactions that were not identified in the previous QSAR study, involving the methyl amino ($N_{4'}$) and the methoxy group ($O_{3'}$) of staurosporine, could play roles in constraining the distance between the N- and C-terminal lobes to the optimal value.

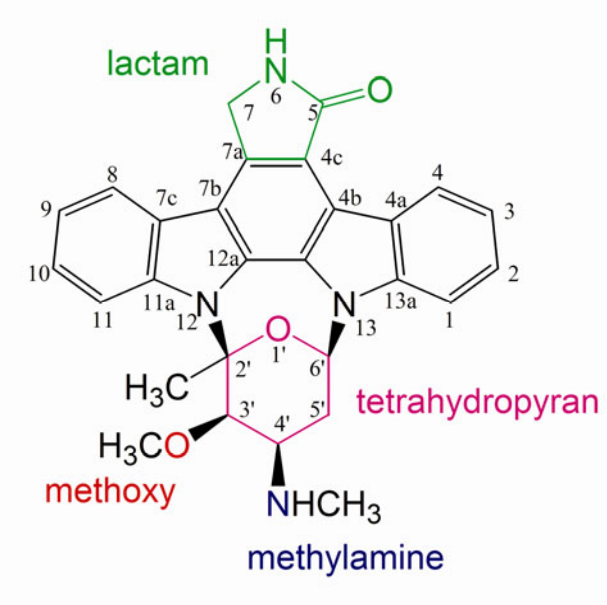


Figure 48. The chemical structure of staurosporine shows the position of N4′ used for MAHORI query.

These distances are neither constant in position so that they can be captured by the grid, nor a linear function of binding affinities that can be captured by multiple linear regression. More distances can be identified by incorporating quadratic terms into the multiple linear regression in the same way as the Hansch equation in Section 1.4. This equation is obtained from collaboration with Professor Amiram Goldblum, who performed the Hansch analysis and suggested that the best way to describe this system is by including the square term in the function. The equation that is obtained using this approach is as follows (Equation 7):

Equation 7.

$$\log K_{d,STU} = 44.82 + 0.02 \times (D_{104\_122})^2 - 0.26 \times D_{120\_122} - 5.89 \times D_{70\_170} + 0.18 \times (D_{70\_170})^2 + 0.25 \times D_{71\_170};$$

where: $K_{d,STU}$ is the dissociation constant of staurosporine, the distance between residues X and Y is written as $D_{X\_Y,}$ and standard deviation (S.D.) = 0.28, $R^2 = 0.83$, R = 0.91.

The $R^2$ is significantly better than the linear equation without the square terms. Although the $R^2$ is higher, there is a trade off between the ability to explain the binding affinities and over-fitting which arises from using a larger number of descriptors. Therefore, I verified the equation with a test set of newly released structures in the PDB and the binding affinities of these structures were found to be predicted with high accuracy (Figure 49).

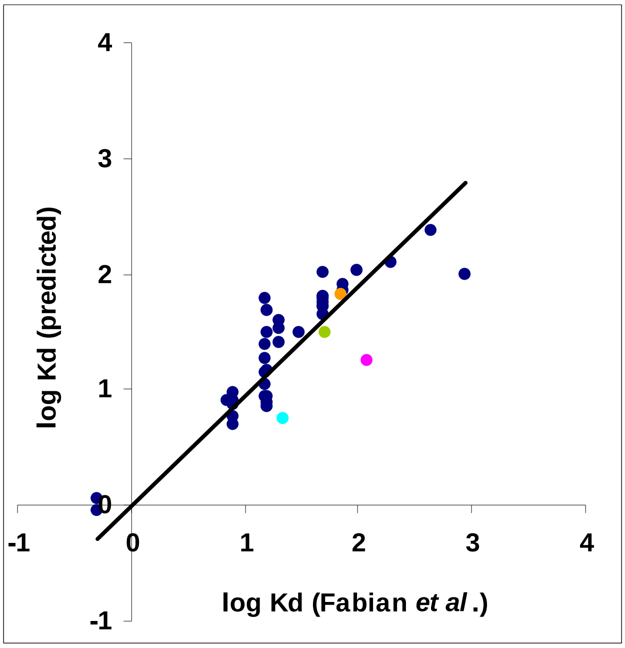


Figure 49. Multiple linear regression with the square distance terms (see equation 7). The navy-dots comprise the dataset from 38 kinases that are used to construct the equation. The non-navy dots are the test set of newly released structures from the Protein Data Bank (cyan, PDB ID 2hw7; green, PDB ID 2dq7; orange, PDB ID 2itq; magenta, PDB ID 2clq).

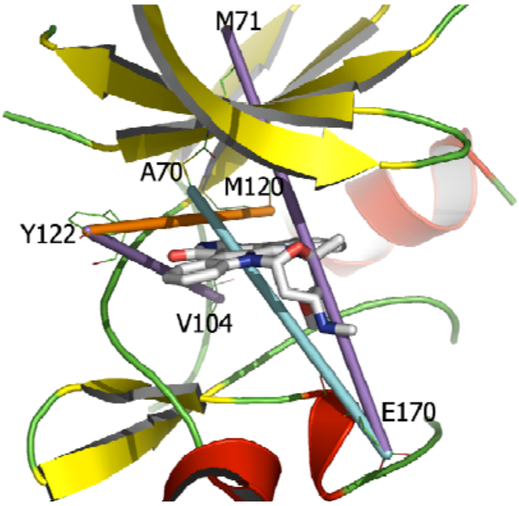


Figure 50. The terms obtained from Equation 7 confirm the role of the size of the gatekeeper ($D_{120\_122}$) and the closure between the N-terminal and C-terminal lobe ($D_{71\_170}$). This equation also suggests an optimal distance along the direction between Ala 70 and Glu 170.

Equation 7 includes two distances in directions that were not present in the equations without square terms in Table 6. The first term is the distance $D_{104\_122}$ (purple bar between Y122 and V104 in Figure 50), linking residues that are found in the tight binding pocket and which should move closer together to act as tweezers that tighten the pocket around the lactam area. The second is the distance $D_{70\_170}$ (blue bar between Ala70 and Glu170 in Figure 50) which is a quadratic term in equation 7. This term implies that an optimal distance is required along the direction of the blue bar. The term $D_{70\_170}$ suggests a role for the distance along the direction that passes through the methylamine $N_{4'}$ of staurosporine (Figure 50).

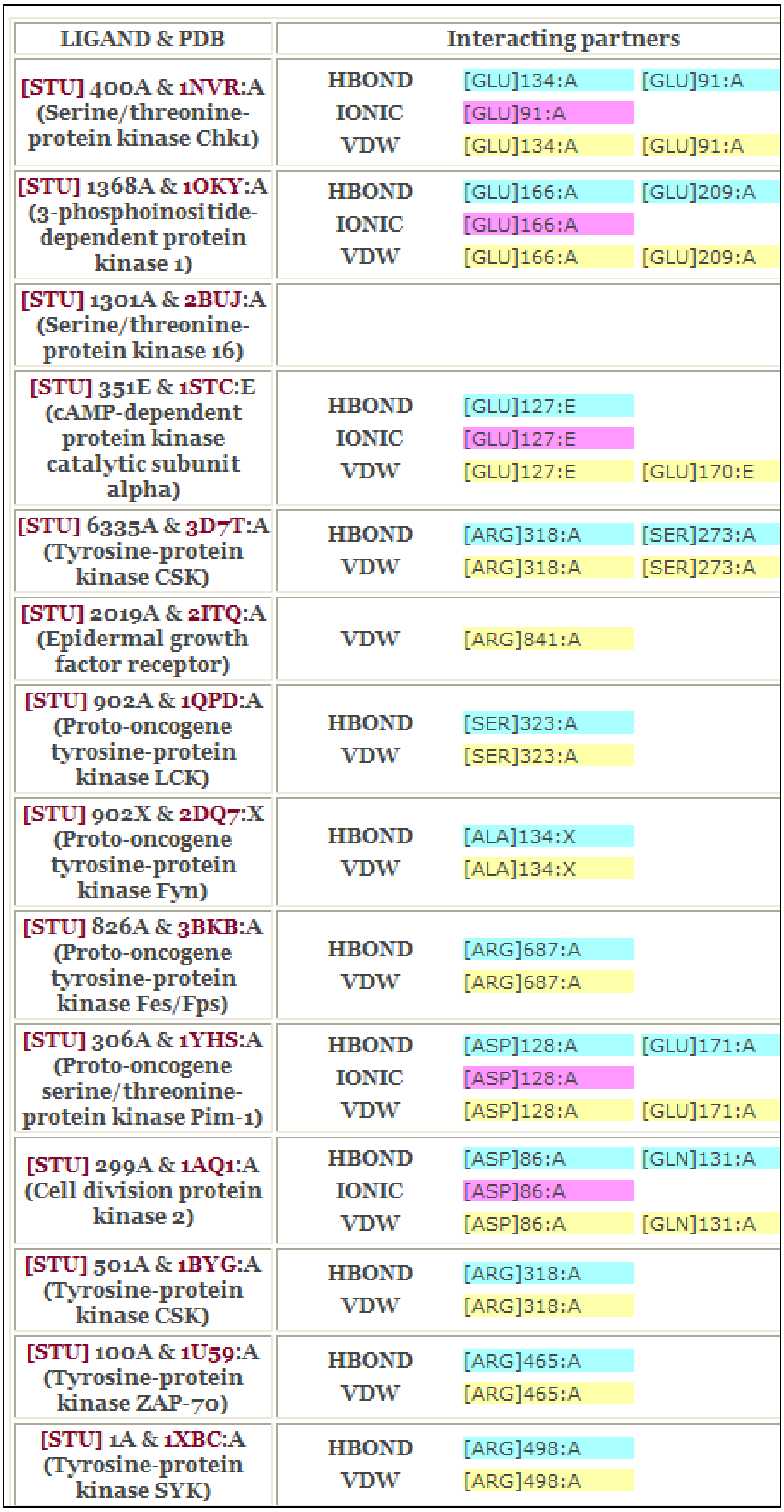
Indeed, the hydrogen bonds or ionic interactions that staurosporine can make along this direction are associated with the major differences in the binding affinities.

I found that the number of hydrogen bonds made by residues around $N_{4'}$ of staurosporine corresponds well with the trend in binding affinity (Table 7). Kinase structures that have two residues making hydrogen bonds or ionic interactions to $N_{4'}$ of staurosporine, *i.e.* CDK2, PKA, PIM1, and LCK, have binding affinities below 51 nM. Most structures that have only one residue contributing hydrogen bonds or ionic interactions to $N_{4'}$ have binding affinities between 51-440 nM, *i.e.* CSK, EGFR, FYN, M3K5. The kinase STK16, which does not make any interaction with $N_{4'}$, has a binding affinity of 200 nM.

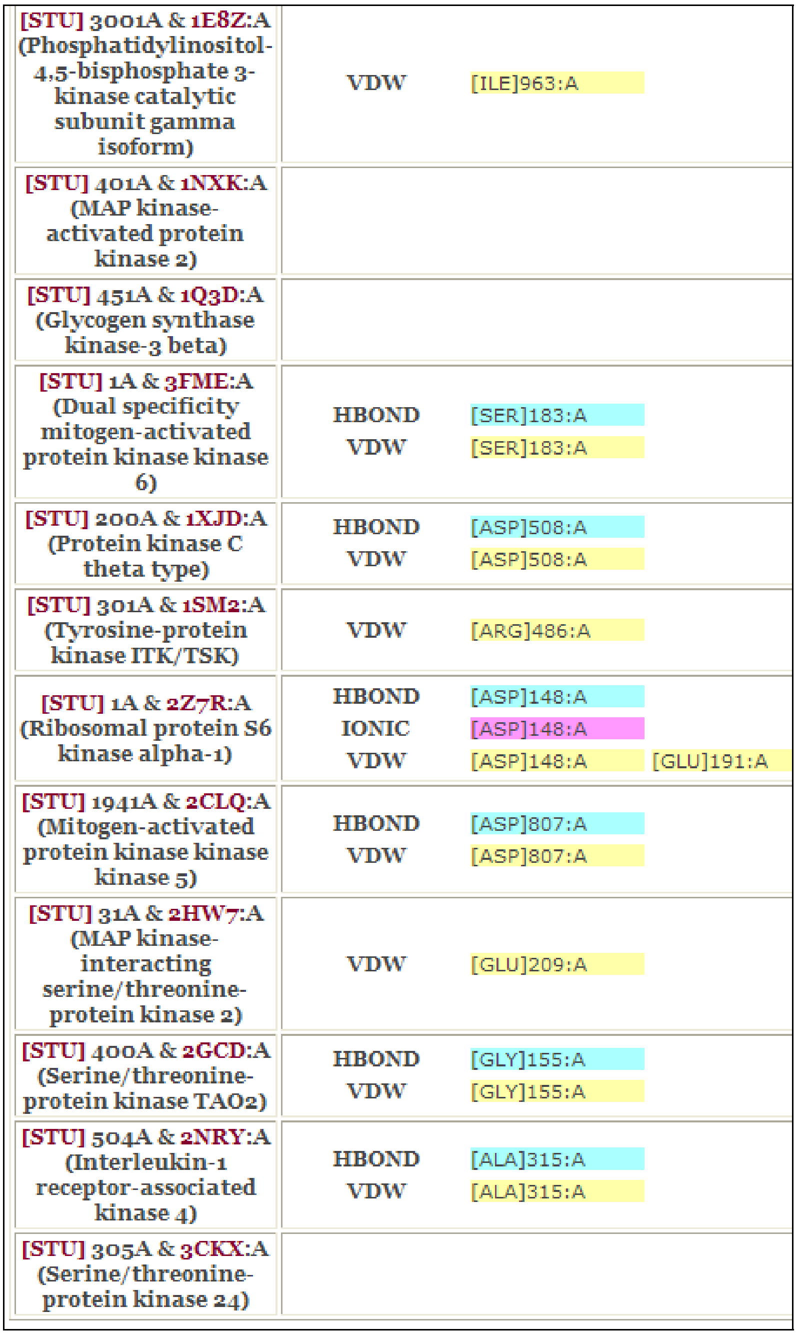
Table 7. Number of interactions made by the kinases with methylamine (N4′) of staurosporine.

| Protein kinase | PDB ID | $K_{d,\text{ STU}}$ (nM) | Number of interactions | | |
|---|---|---|---|---|---|
| | | | H-bond | Ionic | vdW |
| CDK2 | 1AQ1 | 8.1 | 2 | 1 | 2 |
| PIM1 | 1YHS | 15 | 2 | 1 | 2 |
| LCK | 1QPJ | 20 | 2 | - | 2 |
| PKA | 1STC | 50 | 1 | 1 | 2 |
| SYK | 1XBC | 7 | 1 | - | 1 |
| FYN | 2DQ7 | 51 | 1 | - | 1 |
| M3K5 | 2CLQ | 120 | 1 | - | 1 |
| CSK | 1BYG | 440 | 1 | - | 1 |
| MKNK2 | 2HW7 | 22 | - | - | 1 |
| EGFR | 2ITU | 70 | - | 1 | 2 |
| STK16 | 2BUJ | 200 | - | - | - |

In order to achieve better affinity for kinases, the strategy might be to identify a residue in the protein close to the methyl amino ($N_{4'}$) or to modify the staurosporine ligand so that it can make a further hydrogen bond. The reference PKA structure is the PDB ID: 1STC from Figure 51. The majority of kinases make one hydrogen bond with the atom that is equivalent to the main chain carbonyl oxygen of residue 170. When there are two hydrogen bonds made with $N_{4'}$, the additional hydrogen bond comes from the side chain of Glu 127 (Figure 51).



| LIGAND & PDB | Interacting partners | | |
|---|---|---|---|
| [STU] 400A & 1NVR:A (Serine/threonine-protein kinase Chk1) | HBOND | [GLU]134:A | [GLU]91:A |
| | IONIC | [GLU]91:A | |
| | VDW | [GLU]134:A | [GLU]91:A |
| [STU] 1368A & 1OKY:A (3-phosphoinositide-dependent protein kinase 1) | HBOND | [GLU]166:A | [GLU]209:A |
| | IONIC | [GLU]166:A | |
| | VDW | [GLU]166:A | [GLU]209:A |
| [STU] 1301A & 2BUJ:A (Serine/threonine-protein kinase 16) | | | |
| [STU] 351E & 1STC:E (cAMP-dependent protein kinase catalytic subunit alpha) | HBOND | [GLU]127:E | |
| | IONIC | [GLU]127:E | |
| | VDW | [GLU]127:E | [GLU]170:E |
| [STU] 6335A & 3D7T:A (Tyrosine-protein kinase CSK) | HBOND | [ARG]318:A | [SER]273:A |
| | VDW | [ARG]318:A | [SER]273:A |
| [STU] 2019A & 2ITQ:A (Epidermal growth factor receptor) | VDW | [ARG]841:A | |
| [STU] 902A & 1QPD:A (Proto-oncogene tyrosine-protein kinase LCK) | HBOND | [SER]323:A | |
| | VDW | [SER]323:A | |
| [STU] 902X & 2DQ7:X (Proto-oncogene tyrosine-protein kinase Fyn) | HBOND | [ALA]134:X | |
| | VDW | [ALA]134:X | |
| [STU] 826A & 3BKB:A (Proto-oncogene tyrosine-protein kinase Fes/Fps) | HBOND | [ARG]687:A | |
| | VDW | [ARG]687:A | |
| [STU] 306A & 1YHS:A (Proto-oncogene serine/threonine-protein kinase Pim-1) | HBOND | [ASP]128:A | [GLU]171:A |
| | IONIC | [ASP]128:A | |
| | VDW | [ASP]128:A | [GLU]171:A |
| [STU] 299A & 1AQ1:A (Cell division protein kinase 2) | HBOND | [ASP]86:A | [GLN]131:A |
| | IONIC | [ASP]86:A | |
| | VDW | [ASP]86:A | [GLN]131:A |
| [STU] 501A & 1BYG:A (Tyrosine-protein kinase CSK) | HBOND | [ARG]318:A | |
| | VDW | [ARG]318:A | |
| [STU] 100A & 1U59:A (Tyrosine-protein kinase ZAP-70) | HBOND | [ARG]465:A | |
| | VDW | [ARG]465:A | |
| [STU] 1A & 1XBC:A (Tyrosine-protein kinase SYK) | HBOND | [ARG]498:A | |
| | VDW | [ARG]498:A | |



| | | | |
|---|---|---|---|
| [STU] 3001A & 1E8Z:A (Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit gamma isoform) | VDW | [ILE]963:A | |
| [STU] 401A & 1NXK:A (MAP kinase-activated protein kinase 2) | | | |
| [STU] 451A & 1Q3D:A (Glycogen synthase kinase-3 beta) | | | |
| [STU] 1A & 3FME:A (Dual specificity mitogen-activated protein kinase kinase 6) | HBOND | [SER]183:A | |
| | VDW | [SER]183:A | |
| [STU] 200A & 1XJD:A (Protein kinase C theta type) | HBOND | [ASP]508:A | |
| | VDW | [ASP]508:A | |
| [STU] 301A & 1SM2:A (Tyrosine-protein kinase ITK/TSK) | VDW | [ARG]486:A | |
| [STU] 1A & 2Z7R:A (Ribosomal protein S6 kinase alpha-1) | HBOND | [ASP]148:A | |
| | IONIC | [ASP]148:A | |
| | VDW | [ASP]148:A | [GLU]191:A |
| [STU] 1941A & 2CLQ:A (Mitogen-activated protein kinase kinase kinase 5) | HBOND | [ASP]807:A | |
| | VDW | [ASP]807:A | |
| [STU] 31A & 2HW7:A (MAP kinase-interacting serine/threonine-protein kinase 2) | VDW | [GLU]209:A | |
| [STU] 400A & 2GCD:A (Serine/threonine-protein kinase TAO2) | HBOND | [GLY]155:A | |
| | VDW | [GLY]155:A | |
| [STU] 504A & 2NRY:A (Interleukin-1 receptor-associated kinase 4) | HBOND | [ALA]315:A | |
| | VDW | [ALA]315:A | |
| [STU] 305A & 3CKX:A (Serine/threonine-protein kinase 24) | | | |

Figure 51. Interactions made by the atom N4′ of staurosporine in various crystal structures. The PDB code of the protein chain is shown in red and the name of the protein is in brackets. The interacting residues are highlighted with the name of the residue in square brackets followed by the residue number and the protein chain.

Therefore, I speculate that the binding affinities of staurosporine derivatives can be affected most by optimising the interaction with the side chains of residues that are equivalent to Glu 127 in PKA (Asp 86 in CDK2). The type of this residue varies between kinases but the position is aligned well with no insertion or gap in the Baton structural alignment (Figure 26).

Making point mutations of active site residues in order to achieve a better binding affinity to staurosporine might also be achieved by altering Glu 127 of PKA to a residue that can make an optimal hydrogen bond to $N_{4'}$ of staurosporine. This suggestion might be useful for finding a new kinase construct that binds better to staurosporine and may help in producing more proteins that would provide X-ray structures of staurosporine complexes.

### *5.3.2 Understanding simple bioisosteric replacement for synthetic strategy*

According to a study on the bioisosteric similarity of molecules based on structural alignment and observed chemical replacement in drugs (Krier & Hutter, 2009), replacement of variable substructures in staurosporine can be considered in terms of the replacement of $N_{4'}$, $O_{3'}$, and their methyl substituents. Various bioisosteric substitutions can be found as in Table 8. For example, for a methoxy oxygen (OR or O3′ of staurosporine), more exchanges to methylene groups were observed than to fluorine atoms.

From the list in Table 8, the methoxy group can be substituted with NHR or SR, and the amine group can be substituted with OR or =NR to preserve the size and ability to form hydrogen bond of the linker. Choices for methyl group replacement are more diverse but may lead to selectivity for a particular kinase.

Table 8. Most common bioisosteric substitutions regarding one-to-one exchanges of atoms from Krier & Hutter.

| **-OR**<br>$(O_{3'})$ | **-NHR**<br>$(N_{4'})$ | **$-CH_3$** |
|---|---|---|
| $-CH_2-$ | $-CH_2-$ | Cl |
| $CH_3$ | OR | $CH_2R$ |
| NHR | $=CR_2$ | OR |
| $C_{ar}$ | $C_{ar}$ | OH |
| $=CR_2$ | $N_{ar}$ | F |
| $NR_2$ | =NR | $NH_2$ |
| $N_{ar}$ | =NH | Br |
| OH | | I |
| SR | | |
| F | | |

Once the desired kinase is chosen, MAHORI might be used to investigate the tendency of substituting the $N_{4'}$ from the (-NHR) substitution column. In this way the binding affinities for a large set of kinases can be slightly modified and may lead to the design of an inhibitor which is selective for a smaller group of kinases. The possibility of this application depends on the amount of feature interactions available on the Protein Data Bank.

### *5.3.3 Understanding promiscuous substructure of FGFR inhibitors*

I now analyse small molecules that target the FGFR TK in order to understand the similarity of the interactions they make. Four crystal structures of inhibitor-bound FGFR1 TK domains exist in the PDB (PDB IDs: 1AGW, 1FGI, 2FGI and 3C4F). The interactions made with the FGFR were investigated using MAHORI.

Substructure matching using MAHORI demonstrates that particular substructures of the FGFR inhibitor molecules are responsible for kinase promiscuity.

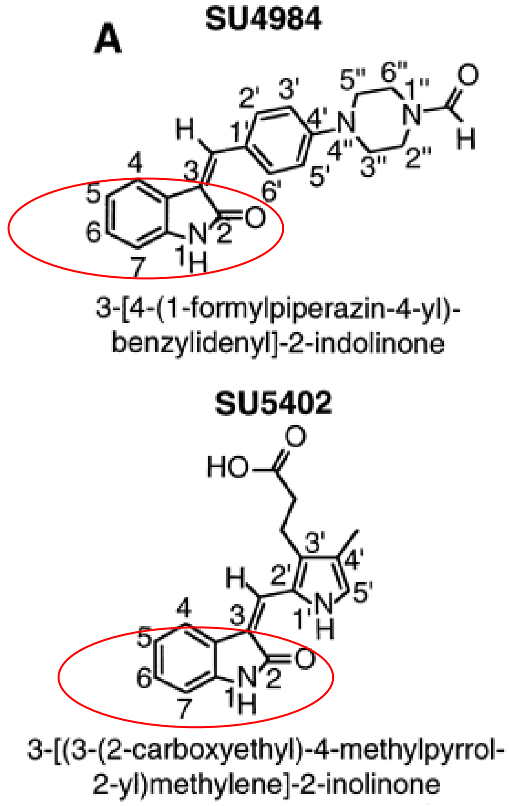

A
SU4984
3-[4-(1-formylpiperazin-4-yl)-benzylidenyl]-2-indolinone
SU5402
3-[(3-(2-carboxyethyl)-4-methylpyrrol-2-yl)methylene]-2-inolinone

Figure 52. The indolinone substructure (red circle) which binds to many kinases.

The indolinone substructure of SU5402 and SU4984 (Mohammadi et al, 1997) interacts with many kinases, such as CDK2 (1PF8, 2BHE, 2BHH, 1KE5-1KE9, 1E9H, 1R78), PIM1 (1YXX), CDK5 (1UNH), GSK3B (1Q41, 1UV5), Casein Kinase 1 homolog1 (1EH4), Serine/Threonine-protein kinase 12-A (2BFY), NEK2 (2JAV), CHK1 (2AYP), and STK10 (2J7T). From the number of kinases that it binds and the position of the indolinone binding area, this substructure has the characteristic of a hinge binder which can be found from the web application MAHORI.

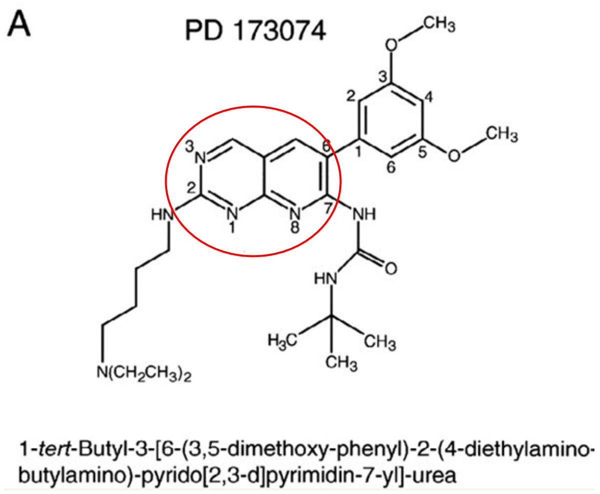

A
PD 173074
CH3
O
3
2
4
CH3
6
1
5
O
6
N
3
2
HN
N
1
N
8
7
NH
O
HN
H3C
CH3
CH3
N(CH2CH3)2
1-*tert*-Butyl-3-[6-(3,5-dimethoxy-phenyl)-2-(4-diethylamino-butylamino)-pyrido[2,3-d]pyrimidin-7-yl]-urea

Figure 53. Pyrido[2,3D]pyrimidinyl substructure (red circle) which binds many kinases as well as many dihydrofolate reductases

The pyrido[2,3D]pyrimidinyl substructure of PD173074 (Mohammadi et al, 1998) also binds ABL1 (1M52, 1OPK, 1OPL, 2FO0, 2G2H), CDK6 (2EUF) and many dihydrofolate reductase structures. The difference is the atom that is equivalent to N3 of the ligand PD173074 makes one hydrogen bond and one aromatic interaction with the hinge region of ABL1, while the same atom makes only an aromatic interaction without hydrogen bonding with a beta-strand in dihydrofolate reductase. This atom does not interact with protein at all in CDK6. This hinge binder substituent may be useful for the design of a multi-target drug, which aims to interfere with both DNA-synthesis via dihydrofolate reductase and signal transduction via kinases.

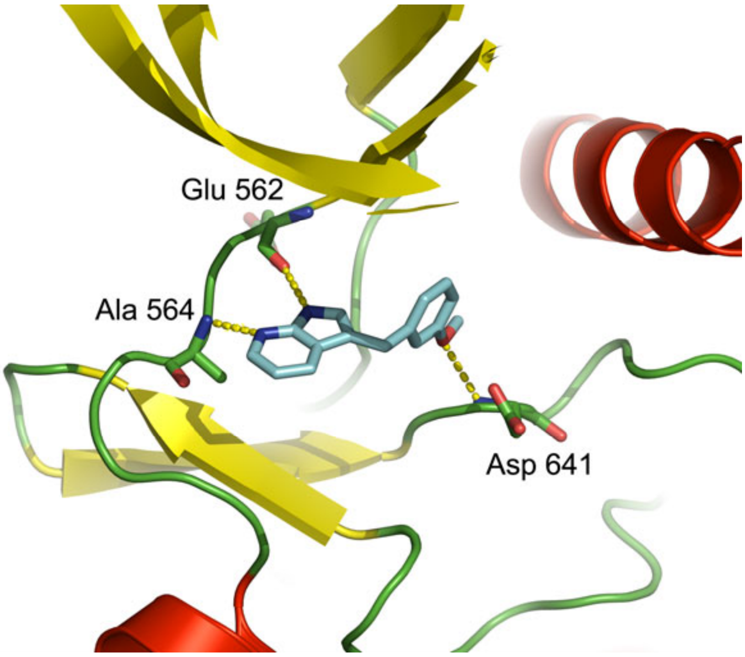


Figure 54. 7-azaindole derivatives in the structure of FGFR from PDB ID 3C4F

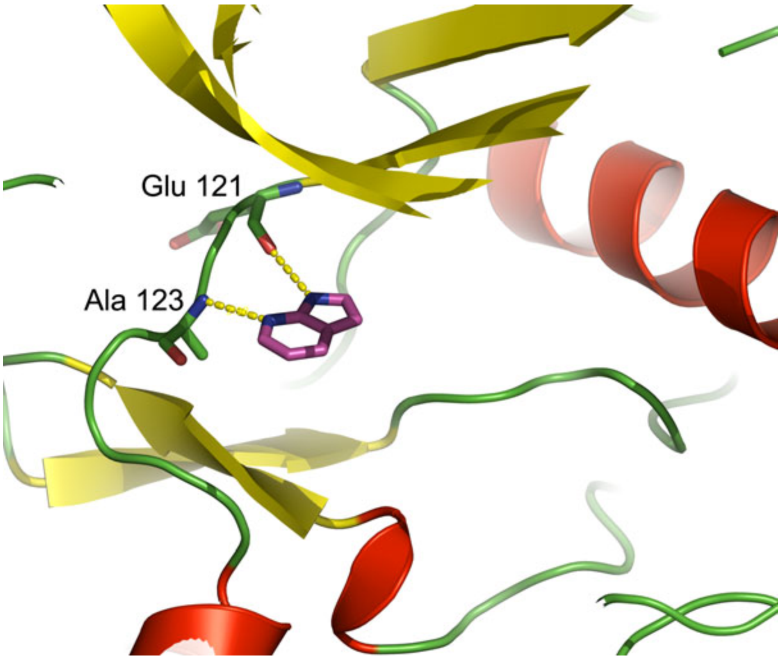


Figure 55. 7-azaindole from PDB ID 2UVX shows its ability to bind cAMP dependent protein kinase without having any decorating substituent, which may be useful as a scaffold for multi-target drug design.

The 1H-pyrrole[2,3-B]pyridine, or 7-azaindole, moiety binds several kinases, such as SRC (3EN4), CDK2 (3BHU) and CHK1 (1ZYS, 2QHM) amongst others. A crystal structure of this fragment alone bound to cyclic-AMP dependent protein kinase (2UVX) implies that 7-azaindole is the promiscuous substructure (Figure 55). This substructure is another hinge binder which may be suitable as a scaffold for multi-target drug design. Decorating substituents should aim to achieve selectivity by limiting the number of kinases to which the molecule can bind.

## 5.4 Conclusion

I have learned that the numbers of hydrogen bonds and ionic interactions determine the magnitudes of binding affinities and the interactions made with methylamine are important to the binding with staurosporine. From this understanding, I can identify Glu 127 as the residue which can be used to achieve better binding affinities for staurosporine derivatives. Promiscuous scaffolds of the ligand can be found in the four examples of FGFR inhibitors which clearly demonstrate the part of the molecules that acts as a hinge binder. I show that three scaffolds which bind to FGFR can also bind to various other kinases. My web-tool, MAHORI, enables comparison of the amount, the type and the position of the interactions in one results page on the computer screen. This kind of information might be useful for rational drug design and medicinal chemistry education.

The tool is useful to confirm the parts of ligands that bind to kinases non-specifically and to rationalise the specificity determining residues. The findings may be helpful in guiding synthetic strategies and mutagenesis studies by narrowing down the choice of chemical group replacement in the ligand and amino acid substitutions in the kinase.

# Chapter 6

## CONCLUDING REMARKS AND FUTURE WORK

At the time of starting the thesis, my research was geared towards the understanding of the determinants of inhibitor selectivity. However, due to the fact that there are so many conformations of kinases with many different kinds of inhibitors available in the Protein Data Bank, my research was refocused on those forms which allow comparisons of many structures; these are staurosporine and ATP complexes.

The problem proved to be not so simple; Q-SiteFinder estimated the active site volumes in CDK2 alone could be hugely variable, for example 182 Å$^3$ (PDB ID 1aq1) with staurosporine and 406 Å$^3$ (PDB ID 1fin) with ATP. This led to the idea of developing an algorithm to observe frequently occurring atoms, because some parts of the pocket that interact with the ligand were held in the same position regardless of their different volumes.

By superposing the rigid parts of the ligands, the staurosporine complexes were proven to be good models. In contrast with ATP binding complexes, staurosporine complexes show a lack of plasticity in the pocket, especially in the constrained region. The superposed position of Gly 50, which varies in ATP complexes but which stays close to oxygen of the tetrahydropyran ring in staurosporine complexes, led to the induced fit hypothesis. The conservation of positions of atoms in the hinge regions for both complexes confirmed that this was the major determinant of cross-reactivity of inhibitors.

Domain superposition provided a new way to compare the rigid parts of the molecule. It became apparent that a residue in the lower middle part of the N-terminal lobe, Ala 70, positions the ligand in protein kinases. This approach could also correctly identify Asp 166 in the C-terminal lobe as a residue that is conserved in position. Interestingly Asp 166 is the amino acid residue which accepts the proton from the substrate and the reason why this region is called the catalytic loop.

The concept of frequently occurring atoms was used further to consider the shape created by these unmoving points. I wanted to know which distance between these points are the most important. Although there is no correlation between distances and binding affinities when the distance between the static atoms is measured, when the distances between the side chains are considered, there is considerable improvement in the correlation with binding affinities. Hence, I developed the idea of measuring the distances between the ends of the side chains.

QSAR indicated that the gatekeeper residue was always involved in the distances that gave rise to the best equations. This coincided with the observation that the size of the gatekeeper residue affects inhibitor binding. Another distance that was important was the one that passes through N4′ of staurosporine. Optimal distances along this direction could result in strong molecular interactions. The greater the number of hydrogen bond interactions and ionic interactions along this direction, the better the binding affinities.

In theory, the MAHORI website could be developed to allow bottom-up fragment design. Indeed, one way to decide whether or not a fragment will stay in a pocket is by considering local environments in other protein crystal structures. Promiscuous scaffolds can also be identified which lead to knowledge about cross-reactivity between enzymes in different pathways. However, at the time of conducting this research, drug-like substructures in the PDB were still quite rare. Hence, there was not much choice in substructure queries using MAHORI. I believe that when the Protein Data Bank grows, the website will become fruitful. There is room for improvement in MAHORI. The first priority is to allow water-mediated interactions. Then I would wish to develop a method to classify the environment and compile a list of frequent substructure interactions. It would be useful if the user could link substructures together to form a new ligand for a particular input pocket.

My idea for future work on kinases is to develop further software for automated identification of specificity determining residues. I would define specificity as the features that are retained after subtracting the promiscuous ones from all those in a particular kinase. Therefore, observing promiscuous features by frequently occurring atoms can be a starting point for obtaining the specificity determining features. Identifying these residues would result in a clear benefit for kinase inhibitor design.

# APPENDIX

## A. Details of protein kinase chains used for shape comparison.

| | Protein | PDB | Length | Res(Å) | R | Species | Ligand | Others | $K_{d,STU}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | **ACK1** | **1U54B** | 291 | 2.80 | 0.23 | HUMAN | **ACP** | MG,PTR | |
| 2 | **AKT2** | **1O6KA** | 336 | 1.70 | 0.20 | HUMAN | **ANP** | MN,TPO | |
| 3 | **CDK2** | **1AQ10** | 298 | 2.00 | 0.22 | HUMAN | **STU** | | 8.1 |
| | **CDK2** | **1B38A** | 299 | 2.00 | 0.18 | HUMAN | **ATP** | MG | 8.1 |
| | **CDK2** | **1B39A** | 299 | 2.10 | 0.20 | HUMAN | **ATP** | MG | 8.1 |
| | **CDK2** | **1FINC** | 298 | 2.30 | 0.21 | HUMAN | **ATP** | | 8.1 |
| | **CDK2** | **1GY3C** | 299 | 2.70 | 0.25 | HUMAN | **ATP** | MG,TPO | 8.1 |
| | **CDK2** | **1JSTC** | 298 | 2.60 | 0.20 | HUMAN | **ATP** | MN,TPO | 8.1 |
| | **CDK2** | **1QMZA** | 299 | 2.20 | 0.22 | HUMAN | **ATP** | MG,TPO | 8.1 |
| 4 | **CHK1** | **1NVRA** | 289 | 1.80 | 0.19 | HUMAN | **STU** | | 30.0 |
| 5 | **CK2A** | **1DAWA** | 327 | 2.20 | 0.22 | MAIZE | **ANP** | MG | |
| 6 | **CSK** | **1BYGA** | 278 | 2.40 | 0.20 | HUMAN | **STU** | | 440.0 |
| 7 | **CSK21** | **1PJKA** | 334 | 2.50 | 0.19 | HUMAN | **ANP** | | |
| | **CSK21** | **1YMIA** | 334 | 1.66 | 0.19 | HUMAN | **ANP** | | |
| 8 | **CSK2A** | **1LP4A** | 332 | 1.86 | 0.21 | MAIZE | **ANP** | MG | |
| 9 | **DAPK1** | **1IG1A** | 294 | 1.80 | 0.19 | HUMAN | **ANP** | MN | |
| | **DAPK1** | **1JKKA** | 294 | 2.40 | 0.20 | HUMAN | **ANP** | MG | |
| | **DAPK1** | **1JKLA** | 294 | 1.62 | 0.20 | HUMAN | **ANP** | | |
| 10 | **EPHA2** | **1MQBA** | 333 | 2.30 | 0.24 | HUMAN | **ANP** | | 870.0 |
| 11 | **EPHB2** | **1JPAA** | 312 | 1.91 | 0.23 | MOUSE | **ANP** | | |
| 12 | **FAK1** | **1MP8A** | 281 | 1.60 | 0.19 | HUMAN | **ADP** | | |
| 13 | **FUS3** | **2B9FA** | 353 | 1.80 | 0.21 | YEAST | **ADP** | MG | |
| | **FUS3** | **2B9IA** | 353 | 2.50 | 0.20 | YEAST | **ADP** | MG | |
| | **FUS3** | **2B9JA** | 353 | 2.30 | 0.20 | YEAST | **ADP** | MG | |
| | **FUS3** | **2B9HA** | 353 | 1.55 | 0.19 | YEAST | **ADP** | MG,STE7 | |
| 14 | **GCN2** | **1ZY5A** | 303 | 2.00 | 0.23 | YEAST | **ANP** | MG | |
| | **GCN2** | **1ZYDA** | 303 | 2.75 | 0.21 | YEAST | **ATP** | MG | |
| 15 | **GSK3B** | **1J1BA** | 420 | 1.80 | 0.22 | HUMAN | **ANP** | | |
| | **GSK3B** | **1J1CA** | 420 | 2.10 | 0.22 | HUMAN | **ADP** | MG | |
| | **GSK3B** | **1PYXB** | 422 | 2.40 | 0.21 | HUMAN | **ANP** | MG | |
| | **GSK3B** | **1Q3DA** | 424 | 2.20 | 0.23 | HUMAN | **STU** | | |
| 16 | **IF2A** | **2A19B** | 284 | 2.50 | 0.23 | YEAST | **ANP** | MG | |
| 17 | **IGF1R** | **1JQHA** | 308 | 2.10 | 0.20 | HUMAN | **ANP** | MG | |
| | **IGF1R** | **1K3AA** | 299 | 2.10 | 0.21 | HUMAN | **ACP** | PTR | |
| 18 | **INSR** | **1I44A** | 306 | 2.40 | 0.21 | HUMAN | **ACP** | MG | 73.0 |
| | **INSR** | **1IR3A** | 306 | 1.90 | 0.19 | HUMAN | **ANP** | MG,PTR | 73.0 |
| 19 | **ITK** | **1SM2A** | 264 | 2.30 | 0.25 | HUMAN | **STU** | - | |
| 20 | **KAPCA** | **1ATPE** | 350 | 2.20 | 0.18 | MOUSE | **ATP** | MN | 50.0 |
| | **KAPCA** | **1BKXA** | 350 | 2.60 | 0.22 | MOUSE | **ADE** | TPO | 50.0 |
| | **KAPCA** | **1FMOE** | 350 | 2.20 | 0.18 | MOUSE | **ADN** | TPO | 50.0 |
| | **KAPCA** | **1JBPE** | 350 | 2.20 | 0.17 | MOUSE | **ADP** | TPO | 50.0 |
| | **KAPCA** | **1L3RE** | 350 | 2.00 | 0.20 | MOUSE | **ADP** | MG,TPO | 50.0 |
| | **KAPCA** | **1Q24A** | 350 | 2.60 | 0.20 | BOVINE | **ATP** | MG,TPO | 50.0 |
| | **KAPCA** | **1RDQE** | 350 | 1.26 | 0.13 | MOUSE | **ADP,ATP** | MG,TPO | 50.0 |
| | **KAPCA** | **1STCE** | 350 | 2.30 | 0.21 | BOVINE | **STU** | | 50.0 |
| | **KAPCA** | **1U7EA** | 350 | 2.00 | 0.17 | MOUSE | **ANP** | MN,TPO | 50.0 |
| 21 | **KIT** | **1PKGA** | 329 | 2.90 | 0.23 | HUMAN | **ADP** | MG,PTR | 100.0 |
| 22 | **KPCT** | **1XJDA** | 345 | 2.00 | 0.20 | HUMAN | **STU** | | |
| 23 | **KSYK** | **1XBCA** | 291 | 2.00 | 0.25 | HUMAN | **STU** | | 7.0 |
| 24 | **LCK** | **1QPCA** | 279 | 1.60 | 0.20 | HUMAN | **ANP** | PTR | 20.0 |
| | **LCK** | **1QPDA** | 279 | 2.00 | 0.20 | HUMAN | **STU** | | 20.0 |
| | **LCK** | **1QPJA** | 279 | 2.20 | 0.21 | HUMAN | **STU** | | 20.0 |
| 25 | **MAPK2** | **1NXKA** | 400 | 2.70 | 0.24 | HUMAN | **STU** | | |
| | **MAPK2** | **1NY3A** | 400 | 3.00 | 0.27 | HUMAN | **ADP** | | |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **26** | **MP2K1** | **1S9JA** | 341 | 2.40 | 0.25 | HUMAN | **ATP** | MG | |
| **27** | **PDPK1** | **1H1WA** | 289 | 2.00 | 0.20 | HUMAN | **ATP** | | |
| | **PDPK1** | **1OKYA** | 310 | 2.30 | 0.22 | HUMAN | **STU** | | |
| | **PDPK1** | **2BIYA** | 310 | 1.95 | 0.19 | HUMAN | **ATP** | | |
| **28** | **PHKG1** | **1QL6A** | 298 | 2.40 | 0.24 | RABBIT | **ATP** | MN | 0.5 |
| | **PHKG1** | **2PHKA** | 277 | 2.60 | 0.25 | RABBIT | **ATP** | MN | 0.5 |
| **29** | **PIM1** | **1XR1A** | 300 | 2.10 | 0.27 | HUMAN | **ANP** | MG | 15.0 |
| | **PIM1** | **1YHSA** | 273 | 2.15 | 0.23 | HUMAN | **STU** | | 15.0 |
| | **PIM1** | **1YI4A** | 273 | 2.40 | 0.21 | HUMAN | **ADN** | | 15.0 |
| | **PIM1** | **1YXTA** | 293 | 2.00 | 0.18 | HUMAN | **ANP** | | 15.0 |
| | **PIM1** | **1YXUA** | 293 | 2.24 | 0.23 | HUMAN | **AMP** | | 15.0 |
| | **PIM1** | **2BZKB** | 313 | 2.45 | 0.19 | HUMAN | **ANP** | | 15.0 |
| **30** | **PKNB** | **1O6YA** | 299 | 2.20 | 0.19 | MYCTU | **ACP** | MG | |
| **31** | **Q9JLS3** | **1U5RA** | 348 | 2.10 | 0.21 | RAT | **ATP** | MG | |
| **32** | **SKY1** | **1Q8YA** | 373 | 2.05 | 0.21 | YEAST | **ADE,ADP** | MG | |
| | **SKY1** | **1Q97A** | 373 | 2.30 | 0.21 | YEAST | **ATP** | ADN,MG | |
| | **SKY1** | **1Q99A** | 373 | 2.11 | 0.22 | YEAST | **ANP** | | |
| **33** | **SRPK1** | **1WBPA** | 397 | 2.40 | 0.23 | HUMAN | **ADP** | | |
| **34** | **STK16** | **2BUJA** | 317 | 2.60 | 0.19 | HUMAN | **STU** | | 200.0 |
| **35** | **STK6** | **1MQ4A** | 272 | 1.90 | 0.23 | HUMAN | **ADP** | MG | 16.0 |
| | **STK6** | **1MUOA** | 297 | 2.90 | 0.26 | HUMAN | **ADN** | | 16.0 |
| | **STK6** | **1OL5A** | 282 | 2.50 | 0.19 | HUMAN | **ADP** | MG,TPO | 16.0 |
| | **STK6** | **1OL6A** | 282 | 3.00 | 0.28 | HUMAN | **ATP** | | 16.0 |
| | **STK6** | **1OL7A** | 282 | 2.75 | 0.26 | HUMAN | **ADP** | MG,TPO | 16.0 |
| | **STK6** | **2C6DA** | 275 | 2.20 | 0.23 | HUMAN | **ANP** | | 16.0 |
| **36** | **ZAP70** | **1U59A** | 287 | 2.30 | 0.22 | HUMAN | **STU** | | |

## B. Published work

Tanramluk D, Schreyer A, Pitt WR, Blundell TL (2009) **On the Origins of Enzyme Inhibitor Selectivity and Promiscuity: A Case Study of Protein Kinase Binding to Staurosporine**. *Chemical Biology & Drug Design* **74:** 16-24.

# REFERENCES

Apsel B, Blair JA, Gonzalez B, Nazif TM, Feldman ME, Aizenstein B, Hoffman R, Williams RL, Shokat KM, Knight ZA (2008) Targeted polypharmacology: discovery of dual inhibitors of tyrosine and phosphoinositide kinases. *Nat Chem Biol* **4:** 691-699

Ballester PJ, Richards WG (2007) Ultrafast shape recognition for similarity search in molecular databases. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science* **463:** 1307-1321

Banck M, Hutchison G, Morley C. (2007) The Open Babel Package.

Bemis GW, Murcko MA (1996) The properties of known drugs. 1. Molecular frameworks. *J Med Chem* **39:** 2887-2893

Bergner A, Gunther J, Hendlich M, Klebe G, Verdonk M (2001) Use of Relibase for retrieving complex three-dimensional interaction patterns including crystallographic packing effects. *Biopolymers* **61:** 99-110

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucl Acids Res* **28:** 235-242

Blume-Jensen P, Hunter T (2001) Oncogenic kinase signalling. *Nature* **411:** 355-365

Böhm H-J, Flohr A, Stahl M (2004) Scaffold Hopping. *Drug Discov Today Tech* **1:** 217-224

Bonnet E, Peer YVd (2002) zt: a software tool for simple and partial Mantel tests. *Journal of Statistical Software* **7:** 1-12

Breitenlechner CB, Bossemeyer D, Engh RA (2005) Crystallography for protein kinase drug design: PKA and SRC case studies. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics*
*Inhibitors of Protein Kinases (4th International Conference, Inhibitors of Protein Kinases) and Associated Workshop: Modelling of Specific Molecular Recognition Processes (Warsaw, Poland, June 25-29, 2005)* **1754:** 38-49

Bron C, Kerbosch J (1973) Algorithm 457: finding all cliques of an undirected graph. *Commun ACM* **16:** 575-577

Bruno IJ, Cole JC, Lommerse JPM, Rowland RS, Taylor R, Verdonk ML (1997) IsoStar: A library of information about nonbonded interactions. *Journal of Computer-Aided Molecular Design* **11:** 525-537

Cheek S, Ginalski K, Zhang H, Grishin NV (2005) A comprehensive update of the sequence and structure classification of kinases. *BMC Struct Biol* **5:** 6

ChemAxonLtd. (2007) MarvinSketch. Budapest.

Cohen P (2002) Protein kinases--the major drug targets of the twenty-first century? *Nat Rev Drug Discov* **1:** 309-315

Connolly M (1983) Analytical molecular surface calculation. *Journal of Applied Crystallography* **16:** 548-558

Cornell WD, Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J.W. & Kollman, P. A. (1995) A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc* **117:** 5179-5197

Crum-Brown A, Fraser TR (1868) On the connection between chemical constitution and physiological action. Part I. On the physiological actions of the salts of the ammonium bases, derived from strychnia, brucia, thebia, codria, morphia, and nicotia. *Trans Roy Soc Edinburgh* **25:** 151-203

Davies TG, Pratt DJ, Endicott JA, Johnson LN, Noble MEM (2002) Structure-based design of cyclin-dependent kinase inhibitors. *Pharmacology & Therapeutics* **93:** 125-133

Debe DA, Hambly K (2004) Supporting your pipeline with structural knowledge. *Curr Drug Discov* **3:** 15-18

DeLano WL. (2002) The PyMOL Molecular Graphics System. DeLano Scientific, Palo Alto, CA, USA.

Ertl P (2003) Cheminformatics analysis of organic substituents: identification of the most common substituents, calculation of substituent properties, and automatic identification of drug-like bioisosteric groups. *J Chem Inf Comput Sci* **43:** 374-380

Fabian MA, Biggs WH, 3rd, Treiber DK, Atteridge CE, Azimioara MD, Benedetti MG, Carter TA, Ciceri P, Edeen PT, Floyd M, Ford JM, Galvin M, Gerlach JL, Grotzfeld RM, Herrgard S, Insko DE, Insko MA, Lai AG, Lelias JM, Mehta SA, Milanov ZV, Velasco AM, Wodicka LM, Patel HK, Zarrinkar PP, Lockhart DJ (2005) A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat Biotechnol* **23:** 329-336

Fahmy T. (2008) XLSTAT 2008. Addinsoft, New York.

Fattori D (2004) Molecular recognition: the fragment approach in lead generation. *Drug Discov Today* **9:** 229-238

Fedorov O, Marsden B, Pogacic V, Rellos P, Muller S, Bullock AN, Schwaller J, Sundstrom M, Knapp S (2007) A systematic interaction map of validated kinase inhibitors with Ser/Thr kinases
10.1073/pnas.0708800104. *Proceedings of the National Academy of Sciences* **104:** 20523-20528

Felsenstein J. (2004) PHYLIP (Phylogeny Inference Package). Department of Genome Sciences and Department of Biology, University of Washington, Seattle.

Force T, Kuida K, Namchuk M, Parang K, Kyriakis JM (2004) Inhibitors of protein kinase signaling pathways: emerging therapies for cardiovascular disease. *Circulation* **109:** 1196-1205

Frantz S (2005) Drug discovery: Playing dirty. *Nature* **437:** 942-943

Free SM, Jr., Wilson JW (1964) A Mathematical Contribution to Structure-Activity Studies. *J Med Chem* **7:** 395-399

Goldstein DM, Gray NS, Zarrinkar PP (2008) High-throughput kinase profiling as a platform for drug discovery. *Nat Rev Drug Discov* **7:** 391-397

Golovin A, Dimitropoulos D, Oldfield T, Rachedi A, Henrick K (2005) MSDsite: a database search and retrieval system for the analysis and viewing of bound ligands and active sites. *Proteins* **58:** 190-199

Golovin A, Oldfield TJ, Tate JG, Velankar S, Barton GJ, Boutselakis H, Dimitropoulos D, Fillon J, Hussain A, Ionides JMC, John M, Keller PA, Krissinel E, McNeil P, Naim A, Newman R, Pajon A, Pineda J, Rachedi A, Copeland J, Sitnov A, Sobhany S, Suarez-Uruena A, Swaminathan GJ, Tagari M, Tromm S, Vranken W, Henrick K (2004) E-MSD: an integrated data resource for bioinformatics
10.1093/nar/gkh078. *Nucl Acids Res* **32:** D211-216

Hammett L (1935) Some relations between reaction rates and equilibrium constants. *Chem Rev* **17:** 125

Hanks SK, Hunter T (1995) Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J* **9:** 576-596

Hanks SK, Quinn AM, Hunter T (1988) The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* **241:** 42-52

Hansch C (1969) A quantitative approach to biochemical structure-activity relationships. *Acc Chem Res* **2:** 232-239

Hansch C, Fujita T (1964) ρ-σ-π-analysis. A method for the correlation of biological activity and chemical structure. *J Am Chem Soc* **86:** 1616-1626

Hemmer W, McGlone M, Tsigelny I, Taylor SS (1997) Role of the Glycine Triad in the ATP-binding Site of cAMP-dependent Protein Kinase 10.1074/jbc.272.27.16946. *J Biol Chem* **272:** 16946-16954

Hendlich M, Rippmann F, Barnickel G (1997) LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J Mol Graph Model* **15:** 359-363, 389

Hopkins AL, Mason JS, Overington JP (2006) Can we rationally design promiscuous drugs? *Current Opinion in Structural Biology* **16:** 127-136

Hu S-H, Parker MW, Yi Lei J, Wilce MCJ, Benian GM, Kemp BE (1994) Insights into autoregulation from the crystal structure of twitchin kinase. *Nature* **369:** 581-584

Hubbard SR (1997) Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. *Embo J* **16:** 5572-5581

Hubbard SR, Wei L, Hendrickson WA (1994) Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature* **372:** 746-754

Huse M, Kuriyan J (2002) The conformational plasticity of protein kinases. *Cell* **109:** 275-282

Jacoby E (2006) Chemogenomics: drug discovery's panacea? *Mol BioSyst* **2:** 218-220

Karaman MW, Herrgard S, Treiber DK, Gallant P, Atteridge CE, Campbell BT, Chan KW, Ciceri P, Davis MI, Edeen PT, Faraoni R, Floyd M, Hunt JP, Lockhart DJ, Milanov ZV, Morrison MJ, Pallares G, Patel HK, Pritchard S, Wodicka LM, Zarrinkar PP (2008) A quantitative analysis of kinase inhibitor selectivity. **26:** 127-132

Kennewell EA, Willett P, Ducrot P, Luttmann C (2006) Identification of target-specific bioisosteric fragments from ligand-protein crystallographic data. *J Comput Aided Mol Des* **20:** 385-394

Keri G, Orfi L, Eros D, Hegymegi-Barakonyi B, Szantai-Kis C, Horvath Z, Waczek F, Marosfalvi J, Szabadkai I, Pato J, Greff Z, Hafenbradl D, Daub H, Muller G, Klebl B, Ullrich A (2006) Signal Transduction Therapy with Rationally Designed Kinase Inhibitors. *Current Signal Transduction Therapy* **1:** 67-95

Kinnings SL, Jackson RM (2009) Binding Site Similarity Analysis for the Functional Classification of the Protein Kinase Family. *Journal of Chemical Information and Modeling* **49:** 318-329

Kleywegt GJ, Jones TA (1994) Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr D Biol Crystallogr* **50:** 178-185

Knighton DR, Zheng JH, Ten Eyck LF, Ashford VA, Xuong NH, Taylor SS, Sowadski JM (1991) Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* **253:** 407-414

Krier M, Hutter MC (2009) Bioisosteric Similarity of Molecules Based on Structural Alignment and Observed Chemical Replacements in Drugs. *Journal of Chemical Information and Modeling* **49:** 1280-1297

Kubinyi H (1993) *QSAR: Hansch analysis and related approaches*, Vol. 1, New York: VCH Publishers.

Kubinyi H. (1998) Comparative Molecular Field Analysis (CoMFA). In Schleyer PvR, Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer III HF, Schreiner PR (eds.), *The Encyclopedia of Computational Chemistry*. John Wiley & Sons, Chichester, Vol. 1, pp. 448-460.

Kuhn D, Weskamp N, Hullermeier E, Klebe G (2007) Functional classification of protein kinase binding sites using Cavbase. *ChemMedChem* **2:** 1432-1447

Kuhn D, Weskamp N, Schmitt S, Hullermeier E, Klebe G (2006) From the similarity analysis of protein cavities to the functional classification of protein families using cavbase. *J Mol Biol* **359:** 1023-1044

Laskowski RA (1995) SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J Mol Graph* **13:** 323-330, 307-328

Laurie AT, Jackson RM (2005) Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics* **21:** 1908-1916

Lee B, Richards FM (1971) The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* **55:** 379-400

Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23:** 127-128

Lewell XQ, Jones AC, Bruce CL, Harper G, Jones MM, McLay IM, Bradshaw J (2003) Drug rings database with web interface. A tool for identifying alternative chemical rings in lead discovery programs. *J Med Chem* **46:** 3257-3274

Liao JJ-L (2007) Molecular Recognition of Protein Kinase Binding Pockets for Design of Potent and Selective Kinase Inhibitors. *Journal of Medicinal Chemistry* **50:** 409-424

Liu Y, Bishop A, Witucki L, Kraybill B, Shimizu E, Tsien J, Ubersax J, Blethrow J, Morgan DO, Shokat KM (1999) Structural basis for selective inhibition of Src family kinases by PP1. **6:** 671-678

Lopez G, Valencia A, Tress M (2007) FireDB--a database of functionally important residues from proteins of known structure
10.1093/nar/gkl897. *Nucl Acids Res* **35:** D219-223

Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) The protein kinase complement of the human genome. *Science* **298:** 1912-1934

Marcou G, Rognan D (2007) Optimizing Fragment and Scaffold Docking by Use of Molecular Interaction Fingerprints. *J Chem Inf Model* **47:** 195-207

Marsden BD, Knapp S (2008) Doing more than just the structure--structural genomics in kinase drug discovery. *Current Opinion in Chemical Biology* **12:** 40-45

Meyer H (1899) Zur theorie der Alkoholnarkose. Erste mittheilung, welche eigenschaft der anasthetica bedingt ihre narkotiste wirkung. *Arch Exp Pathol Pharmakol* **42:** 109

Mizuguchi K, Deane C, Blundell T, Johnson M, Overington J (1998) JOY: protein sequence-structure representation and analysis 10.1093/bioinformatics/14.7.617. *Bioinformatics* **14:** 617-623

Mohammadi M, Froum S, Hamby JM, Schroeder MC, Panek RL, Lu GH, Eliseenkova AV, Green D, Schlessinger J, Hubbard SR (1998) Crystal structure of an angiogenesis inhibitor bound to the FGF receptor tyrosine kinase domain. *Embo J* **17:** 5896-5904

Mohammadi M, McMahon G, Sun L, Tang C, Hirth P, Yeh BK, Hubbard SR, Schlessinger J (1997) Structures of the Tyrosine Kinase Domain of Fibroblast Growth Factor Receptor in Complex with Inhibitors. *Science* **276:** 955-960

Morgan DO (1997) Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu Rev Cell Dev Biol* **13:** 261-291

Morris RJ, Najmanovich RJ, Kahraman A, Thornton JM (2005) Real spherical harmonic expansion coefficients as 3D shape descriptors for protein binding pocket and ligand comparisons. *Bioinformatics* **21:** 2347-2355

Muller G (2003) Medicinal chemistry of target family-directed masterkeys. *Drug Discov Today* **8:** 681-691

Nayal M, Honig B (2006) On the nature of cavities on protein surfaces: application to the identification of drug-binding sites. *Proteins* **63:** 892-906

Niedner RH, Buzko OV, Haste NM, Taylor A, Gribskov M, Taylor SS (2006) Protein kinase resource: an integrated environment for phosphorylation research. *Proteins* **63:** 78-86

Noble ME, Endicott JA, Johnson LN (2004) Protein kinase inhibitors: insights into drug design from structure. *Science* **303:** 1800-1805

Nolen B, Taylor S, Ghosh G (2004) Regulation of protein kinases; controlling activity through activation segment conformation. *Mol Cell* **15:** 661-675

Nurse PM (2002) Cyclin dependent kinases and cell cycle control. *Bioscience Reports* **22:** 487-499

Overton E (1901) *Studien Uber Die Narkose, Zugleich eion Beitrag zur Allgemeine Pharmakologie*, Jenna: Fisher.

Page RDM (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12:** 357-358

Prade L, Engh RA, Girod A, Kinzel V, Huber R, Bossemeyer D (1997) Staurosporine-induced conformational changes of cAMP-dependent protein kinase catalytic subunit explain inhibitory potential. **5:** 1627-1637

Roth BL, Sheffler DJ, Kroeze WK (2004) Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. *Nat Rev Drug Discov* **3:** 353-359

Sachsenheimer W, Schulz GE (1977) Two conformations of crystalline adenylate kinase. *Journal of Molecular Biology* **114:** 23-36

Sali A, Overington JP, Johnson MS, Blundell TL (1990) From comparisons of protein sequences and structures to protein modelling and design. *Trends Biochem Sci* **15:** 235-240

Scheeff ED, Bourne PE (2005) Structural Evolution of the Protein Kinase-Like Superfamily. *PLoS Comput Biol* **1:** e49

Schindler T, Sicheri F, Pico A, Gazit A, Levitzki A, Kuriyan J (1999) Crystal Structure of Hck in Complex with a Src Family Selective Tyrosine Kinase Inhibitor. **3:** 639-648

Schreyer A, Blundell T (2009) CREDO: A Protein-Ligand Interaction Database for Drug Discovery. *Chemical Biology & Drug Design* **73:** 157-167

Schulze-Gahmen U, Brandsen J, Jones HD, Morgan DO, Meijer L, Vesely J, Kim SH (1995) Multiple modes of ligand recognition: crystal structures of cyclin-dependent protein kinase 2 in complex with ATP and two inhibitors, olomoucine and isopentenyladenine. *Proteins* **22:** 378-391

Sheinerman FB, Giraud E, Laoui A (2005) High affinity targets of protein kinase inhibitors have similar residues at the positions energetically important for binding. *J Mol Biol* **352:** 1134-1156

Sheridan RP (2002) The Most Common Chemical Replacements in Drug-Like Compounds. *J Chem Inf Comput Sci* **42:** 103-108

Smith R. (2006) KINASEMAP. Cambridge.

Stehelin D, Varmus HE, Bishop JM, Vogt PK (1976) DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature* **260:** 170-173

Taylor SS, Yang J, Wu J, Haste NM, Radzio-Andzelm E, Anand G (2004) PKA: a portrait of protein kinase dynamics. *Biochim Biophys Acta* **1697:** 259-269

TheJmolDevelopmentTeam. (2007) Jmol: an open-source Java viewer for chemical structures in 3D.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25:** 4876-4882

Traxler P, Furet P (1999) Strategies toward the Design of Novel and Selective Protein Tyrosine Kinase Inhibitors. *Pharmacology & Therapeutics* **82:** 195-206

Via A, Ferre F, Brannetti B, Helmer-Citterich M (2000) Protein surface similarities: a survey of methods to describe and compare protein surfaces. *Cell Mol Life Sci* **57:** 1970-1977

Wang G, Dunbrack RL, Jr. (2003) PISCES: a protein sequence culling server. *Bioinformatics* **19:** 1589-1591

Weininger D (2002) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences* **28:** 31-36

Wermuth CG (2006) Similarity in drugs: reflections on analogue design. *Drug Discovery Today* **11:** 348-354

wwPDB. (2007) Heterogen Section. *Protein Data Bank Contents Guide: Atomic Coordinate Entry Format Description Version 3.0.1*, Vol. 2007.

Xu W, Doshi A, Lei M, Eck MJ, Harrison SC (1999) Crystal Structures of c-Src Reveal Features of Its Autoinhibitory Mechanism. *Molecular Cell* **3:** 629-638

Zhang X, Crespo A, Fernández A (2008) Turning promiscuous kinase inhibitors into safer drugs. *Trends in Biotechnology* **26:** 295-301

Zhao H (2007) Scaffold selection and scaffold hopping in lead generation: a medicinal chemistry perspective. *Drug Discov Today* **12:** 149-155

Zheng J, Trafny EA, Knighton DR, Xuong N, Taylor SS, Ten Eyck LF, Sowadski JM (1993) 2.2 A refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor. *Acta Crystallographica Section D* **49:** 362-365

Zimmermann GR, Lehár J, Keith CT (2007) Multi-target therapeutics: when the whole is greater than the sum of the parts. *Drug Discovery Today* **12:** 34-42