IDENTIFICATION OF CONSERVED STRUCTURAL MOTIFS ASSOCIATED WITH PHOSPHORYLATION SITES IN KINASES

by

SURABHI MAHESHWARI

(Under the Direction of Natarajan Kannan)

ABSTRACT

Background: Protein phosphorylation plays a crucial role in the regulation of several cellular processes. Protein kinases are enzymes that catalyze the event of phosphorylation and are themselves regulated by phosphorylation. Regulation by phosphorylation involves formation of a key structural motif within kinases called the RD-pocket. The canonical RD-pocket is formed by three basic residues projecting from different regions of the kinase. Although the importance of the motif has been appreciated in some kinase families, whether or not this is a general feature of other kinases is not well understood. Also, recurrence of the motif in other regions of the kinase for regulation has not been well explored. With the goal of using the structural motif as a feature for predicting allosteric sites we performed a detailed analysis of the RD-pocket in kinases.

Results: Using structural bioinformatics tools we performed a detailed analysis of the canonical RD-pocket in a library of 1924 kinase structures. We found that apart from the canonical RD-pocket which is well known to be associated with regulation,

several alternate RD-pockets exist. Study of the alternate RD-pockets suggests that the altered location of the RD-pockets have presumably emerged to meet the specific requirements of every kinase family. In addition to the RD-pockets, several other pockets called the “non-RD” pockets are in found in other regions of the kinase domain. In a library of 1924 kinase structures, 35 canonical RD-pockets, 39 alternate RD-pockets and 57 non-RD pockets are identified with RMSD of 1 Å from canonical RD-pocket in CDK. Detailed analysis of non-RD pockets in kinase families such as SRC, CDK2, MAPK8, VRK2 and CHK1 provided important clues regarding the functional role of the pockets. Specifically in MAPK8, we found the non-RD pocket is associated with the C-terminal tail tyrosine (Y357). Potential phosphorylation of the tyrosine can lead to conformational changes in the kinase domain, thereby leading to regulation or activation.

Conclusion: The analysis suggests that the basic pocket is a characteristic phospho-acceptor site in kinases whose main function is to allosterically couple the phosphorylation site to distal active and regulatory sites. The recurrent pocket is an important structural feature that can be used in the prediction of potential allosteric sites regulated by phosphorylation in kinases.

INDEX WORDS: phosphorylation, protein kinase, RD-pocket, regulation, allostery

IDENTIFICATION OF CONSERVED STRUCTURAL MOTIFS ASSOCIATED WITH PHOSPHORYLATION SITES IN KINASES

by

SURABHI MAHESHWARI

B.E., S.G.S.I.T.S, INDIA 2010

A Thesis Submitted to the Graduate Faculty of the University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2012

© 2012

SURABHI MAHESHWARI

All Rights Reserved

IDENTIFICATION OF CONSERVED STRUCTURAL MOTIFS ASSOCIATED WITH PHOSPHORYLATION SITES IN KINASES

by

SURABHI MAHESHWARI

Major Professor:	Natarajan Kannan
Committee:	Jan Mrazek
	Shaying Zhao

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2012

DEDICATION

This thesis is dedicated to my family for their endless support and encouragement.

ACKNOWLEDGEMENTS

This work would not have been completed without help and support of many individuals. I would like to thank everyone who has helped me along the way. Above all, I would like to thank my principal supervisor, Dr. Natarajan Kannan for providing me an opportunity to conduct my master's research under him. I am grateful for his guidance, support and for keeping me focused in my research. I would also like to express my gratitude to Dr. Jan Mrazek and Dr. Shaying Zhao for serving on my thesis committee and their valuable suggestions. I am particularly thankful for the continued input of Dr. Krishnadev Oruganty for this project. Our conversations and work together have greatly influenced this thesis. Lastly, I would like to extend my gratitude towards my family without whose support none of this would have been possible.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS v

LIST OF TABLES viii

LIST OF FIGURES ix

CHAPTER

1 INTRODUCTION 1

2 RESULTS 9

2.1 Sequence and Structural Conservation of RD-Pockets 9

2.2 Non-RD Pockets 23

2.3 Non-Kinase Basic Pockets 38

3 CONCLUSION 49

4 METHODS 50

4.1 Sequence Data 50

4.2 Multiple Sequence Alignment 50

4.3 Residue Conservation 50

4.4 Structural Conservation 51

4.5 Analysis of Energy of Interactions 51

4.6 Identification of Basic Pockets in Proteins 52

REFERENCES 53

APPENDICES

A Supplementary Table 64

LIST OF TABLES

Page

Table 2.1: Pockets identified by SPASM in full length library 15

Table 2.2: Pockets identified by SPASM in representative library 16

Table 2.3: Distinct RD pockets 19

LIST OF FIGURES

Page

Figure 1.1: Protein Kinase 3

Figure 2.1: Canonical RD pocket from 10

Figure 2.2: Sequence Conservation of the canonical RD-Pocket residues in Kinases.. 10

Figure 2.3: Conservation of Basic Residue at RD-Arg position 11

Figure 2.4: Conservation of Basic Residue at DFG+3 position 12

Figure 2.5: Conservation of Basic Residue at canonical C-helix position 13

Figure 2.6: Conservation of the canonical RD pocket 14

Figure 2.7: Phosphorylation site in Ser/Thr and Tyrosine Kinases 17

Figure 2.8: Diverse RD-Pockets 18

Figure 2.9: Alternate RD Pocket in FGFR 21

Figure 2.10: Alternate RD-pocket in Aurora 22

Figure 2.11: Alternate RD-pocket in PKR 23

Figure 2.12: Non-RD-pocket in SRC 26

Figure 2.13: Non-RD-pocket in CDK2 29

Figure 2.14: Non-RD-pocket in JNK1 32

Figure 2.15: Non RD-pocket in VRK2 34

Figure 2.16: Alternative RD-Pocket in Chk1 37

Figure 2.17: Basic pocket in glycogen phosphorylase 39

Figure 2.18: Basic pocket in Deubiquitinase 41

Figure 2.19: Basic pocket in SMAD proteins 42

Figure 2.20: Comparison of basic pocket in phosphorylated and non-phosphorylated . 44

Figure 2.21: Basic Pocket in CrkII 46

CHAPTER 1

INTRODUCTION

Reversible protein phosphorylation is one of the most important and widespread posttranslational modification that occurs in the cell. It is a ubiquitous mechanism for cellular regulation and affects every basic cellular process, be it signal transduction, transport, metabolism or cellular growth and development. The importance of protein phosphorylation as a regulatory mechanism was established after the comprehensive study of glycogen phosphorylase by Krebs and Fischer in the late 1950 and their follow up discovery of phosphorylase kinase (Fischer & Krebs, 1955; Krebs, 1998)This led to the discovery of a large family of enzymes called protein kinases that catalyze the event of protein phosphorylation.

Protein kinases transfer the gamma phosphate from the energy donor, ATP, to specific residues in target proteins. The eukaryotic protein kinase (ePK) super family has been classified into seven major subfamilies – AGC, CAMK, CMGC, CK1, STE, TK and TKL (Manning, Whyte, Martinez, Hunter, & Sudarsanam, 2002). In eukaryotes, most important sites of phosphorylation are serine, threonine and tyrosine. Phosphorylation of these residues brings about functional changes in the target protein; it can activate or deactivate the protein, cause or prevent disease mechanisms, change its association with other proteins or create a protein recognition site for another protein (O. W. Huang *et al.*, 2012; L N Johnson, Noble, & Owen, 1996; Louise N Johnson & Lewis, 2001; Krebs & Fischer, 1956). Over-expression, misregulation or mutation of

protein kinases is often implicated in diseases, especially cancer (Greenman *et al.*, 2007), and this has made kinases the prime targets for cancer intervention (Louise N Johnson, 2009).

Although protein kinases regulate the function of several target proteins, many kinases are themselves regulated through phosphorylation by other kinases. Phosphorylation of one or more residues on a critical polypeptide segment known as activation loop is a key regulatory element of many protein kinases. Phosphorylation allows the activation segment to refold causing a major conformational change, relieving any steric blockage of the catalytic or substrate binding site and at the same time allows the formation of a platform against which the substrate can be docked (Hubbard, 1997; Hubbard, Wei, Ellis, & Hendrickson, 1994; Huse & Kuriyan, 2002; L N Johnson *et al.*, 1996; Louise N Johnson & Lewis, 2001). In fact, phosphorylation site acts as an organizing centre, inducing the formation of a network of key residue contacts at the active site necessary for phosphoryl transfer (Knighton, Zheng, Ten Eyck, Xuong, *et al.*, 1991; Knighton, Zheng, Ten Eyck, Ashford, *et al.*, 1991). Kinases regulated by the activation loop phosphorylation have a distinctive structural feature. A set of basic residues cluster around the phospho amino-acid and form a pocket inside which the phosphate moiety sits. This pocket, also known as the RD-pocket (Nolen, Taylor, & Ghosh, 2004), is a characteristic of the active state conformation (Figure 1.1) (L N Johnson *et al.*, 1996). The interaction of the RD-pocket with the phospho-amino acid neutralizes the charge and also reorients ATP-binding residues.

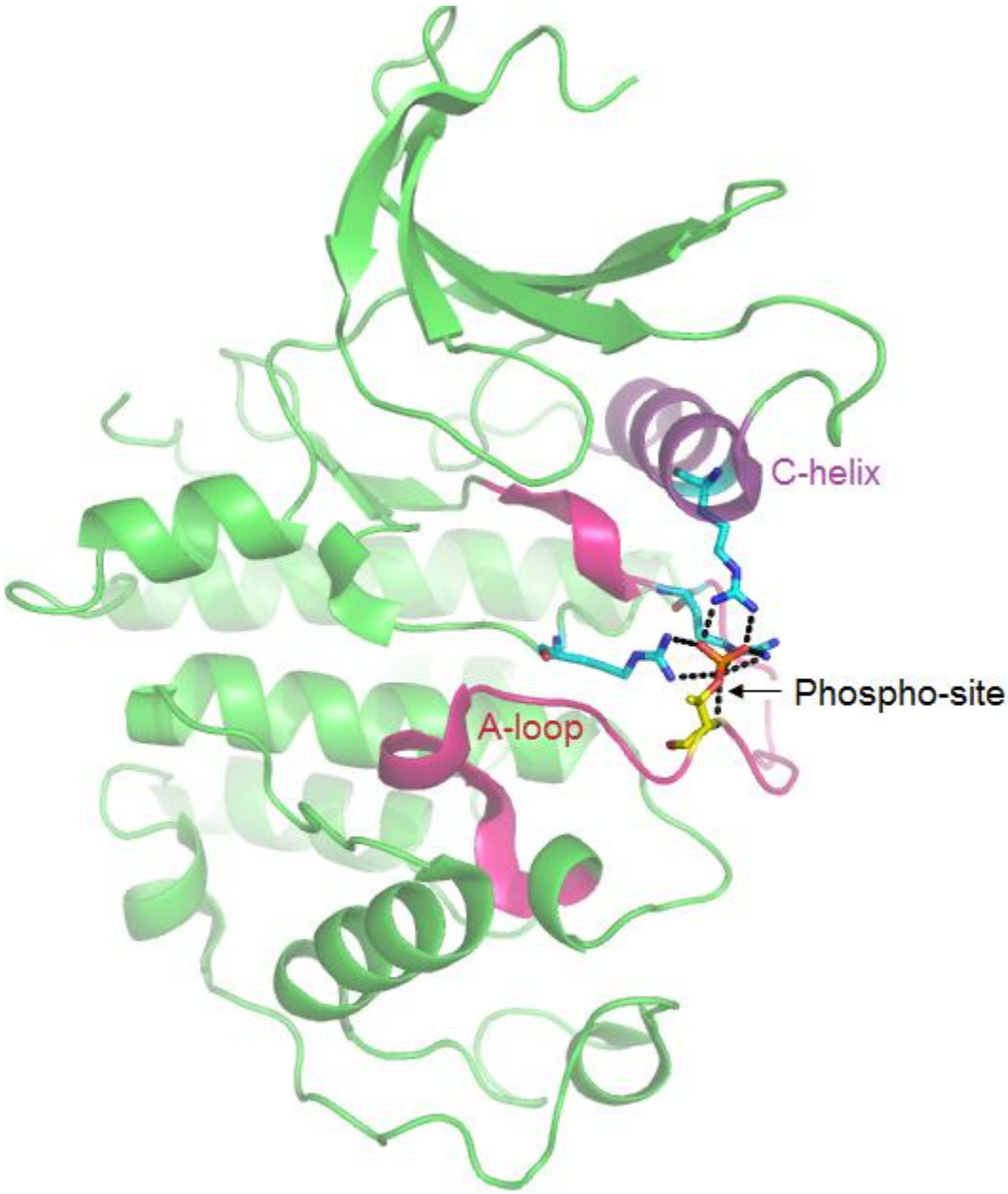


Figure 1.1: Protein Kinase. Overall structure of an active state protein kinase is shown in ribbon diagram (PDB ID 1QMZ). The activation loop is colored pink and C-helix is colored deep purple. The activation loop phospho-site is colored yellow (stick representation). The RD-pocket is shown in stick representation and colored in cyan.

About 30-50% of all of eukaryotic proteins are suggested to undergo reversible phosphorylation (Pinna, Ruzzene, Biologica, & Padova, 1996). Given the importance of phosphorylation in the regulation of most biological events, there has been tremendous effort in the identification of phosphorylation sites in kinases as well as other proteins to better understand their regulatory mechanisms. As a result, increasing number of large scale phosphorylation studies are being conducted. Classical experimental approaches commonly used for identification and validation of phosphorylation sites are mass spectrometry & p-labeling (Schmelzle & White, 2006). These methods may involve onerous procedures and are both expensive and time consuming. Therefore, it is desirable to have efficient methods for predicting phosphorylation sites *in silico*, in order to screen putative sites for experimental validation.

In silico approaches have indeed gained immense popularity and several computational methods have been developed to identify potential phosphorylation sites. Netphos (N Blom, Gammeltoft, & Brunak, 1999), Scansite 2.0 (Obenauer, Cantley, & Yaffe, 2003), KinasePhos 1.0 (H.-D. Huang, Lee, Tzeng, & Horng, 2005), PPSP (Xue, Li, Wang, Feng, & Yao, 2006) are some examples of established computational prediction methods. Phospho-site prediction algorithms can be categorized into two groups. The first group of predictors focuses on anticipating whether the query residue can be phosphorylated or not. Netphos (N Blom *et al.*, 1999) belongs to this category. It was the first phosphorylation site prediction method that identified potential phosphorylation site using consensus-motif-based approaches based on an artificial neural network algorithm. The second category of predictors aims to identify kinase specific phosphorylation sites. Scansite 2.0 (Obenauer *et al.*, 2003), KinasePhos (H.-D.

Huang *et al.*, 2005) & PSSP (Xue *et al.*, 2006) belong to the second category. ScanSite 2.0 (Obenauer *et al.*, 2003) identifies short sequence motifs within proteins that are likely to be phosphorylated by specific protein kinases or bind to phospholipid ligands or domains such as SH2-domain or 14-3-3 domain. KinasePhos (H.-D. Huang *et al.*, 2005) predicts kinase-specific phosphorylation sites within a given protein sequence using computational models built based on hidden Markov Model. Computational models are learned from kinase-specific group of sequences surrounding to the phosphorylation site. PSSP (Xue *et al.*, 2006) uses Bayesian Decision Theory for the prediction of kinase-specific phosphorylation sites.

The above mentioned methods solely rely on the primary sequences around the phosphorylation sites. Although these methods are reasonably sensitive, they often suffer from high false positive rates (Nikolaj Blom, Sicheritz-Pontén, Gupta, Gammeltoft, & Brunak, 2004; Diella *et al.*, 2004). Also, these methods totally neglect some very important structural characteristics of the phosphorylation site: 1) Phosphorylation site can interact directly with amino acids which are in close proximity spatially but not sequentially (as in the case of Protein kinases, three basic residues coming from distinct regions of the kinase domain coordinate with the activation loop phospho-amino acid (Knighton, Zheng, Ten Eyck, Ashford, *et al.*, 1991)) 2) Solvent accessibility & structural disorder indices are characteristic features of phosphorylatable residues (Iakoucheva *et al.*, 2004).

Recently some new algorithms have integrated protein secondary structure information into phosphorylation site prediction. In 2004, Iakoucheva *et al.* hypothesized that phosphorylation generally occurs in intrinsically disordered regions. They

developed DISPHOS, which employs a Support Vector Machine learning algorithm. The algorithm utilizes the structure disorder information along with position-specific residue frequency to improve discrimination between phosphorylation and non-phosphorylation sites (Iakoucheva *et al.*, 2004). In 2005, Plewczynski *et al.* used a molecular modeling approach to delineate the local structural conformation of the phosphorylation sites (Plewczyński, Tkacz, Godzik, & Rychlewski, 2005). They used a sequence matching algorithm to compare protein sequences to a curated library of structural motifs in order to identify potential phosphorylation sites. In 2007, Gnad and collaborators developed PHOSIDA (Gnad *et al.*, 2007), a Support vector machine algorithm which is trained by over 5000 high confidence phosphosites. It extracted the sequence as well as the secondary structure and accessibility information from the training set to predict putative non-specific phospho-sites (Gnad *et al.*, 2007). In 2009, Durek and co-workers characterized the 3D-signature motifs in kinase substrates (Durek, Schudoma, Weckwerth, Selbig, & Walther, 2009). They incorporated 3D-signature motif information with other structural features to improve the prediction of phosophosites (Durek *et al.*, 2009). All the aforementioned methods showed improved predictive power and had lower false positive rates compared to the sequence-only-based methods. These observations further verify that although local sequence surrounding the target amino acid is an important feature for predicting phospho-sites, information regarding the three dimensional environment can substantially improves the prediction.

To identify the specific 3D arrangements of residues surrounding the phospho-site, attempts have been made to explore the 3D environments of phospho-sites. In 2005, Fan and Zhang characterized the microenvironment around the phosphorylated

sites using a simplified “Altman” shell model of radius 16A and compared it with non-phosphorylated sites (Fan & Zhang, 2005). Unlike the enrichment properties identified by sequences-based methods, their analysis highlighted new features around the phospho-sites, such as prevalence of hydrophobic and non polar residues. They concluded that compared to average protein composition there is only a slight aberration in the amino acid composition around the phosphorylation site. Kitchen and coworkers carried out a study that focused on the charged environments around phospho-sites (Kitchen, Saunders, & Warwicker, 2008). They reported that only around one third of the phosphorylated proteins in their dataset were electrostatically stabilized via basic side chain interactions. They suggested that the favorable interactions were evident mainly after conformational change is induced upon phosphorylation. This further clarifies why the “one-fits-all” approach of Fan and Zhang was unable to identify enrichment of basic residues around the phospho-site. Not all phosphorylation events induce a conformational change, and high affinity charged interaction with basic side chain is a feature of phosphorylation sites that induce conformational changes.

Phosphorylation-induced conformational changes are an essential attribute of activation loop phosphorylation in kinases (L N Johnson *et al.*, 1996). As mentioned before, activation loop phosphorylation is a key regulatory event for many protein kinases. Upon phosphorylation, the activation loop undergoes large conformational changes that stabilize the active state conformation of the kinase domain (Nolen *et al.*, 2004). The RD pocket behaves as a phospho-recognition site and tightly binds the phosphate group by establishing hydrogen bonds and ionic interaction with it (L N Johnson *et al.*, 1996). The formation of the RD pocket also reorganizes different regions

of the kinase domain in active state favorable conformation (Nolen *et al.*, 2004). Despite the known functional roles of the RD pocket, however, the RD pocket has not been used as a feature for predicting phospho-sites, or allosteric sites in proteins.

Here, we have performed a systematic analysis of the RD pockets across the kinase super-family, with the ultimate goal of predicting allosteric/phospho-sites in proteins. We observed that although the RD-pocket is not conserved sequentially, it is conserved structurally. We found several alternate RD-pockets where the basic residues emanate from regions different than that of the canonical pocket (RD Arginine being absolutely conserved). Next, we wanted to examine the prevalence of RD-like pockets in other regions of the kinase. We used a spatial motif finder called SPASM (Kleywegt, 1999) to identify such non-RD pockets. We indeed found quite a few non-RD pockets. Analysis of the pockets in light of available literature data provided important clues regarding the functional role of the identified pockets. One such non-RD pocket encloses a very well known negative regulatory phosphosite in the C-terminal tail of Src tyrosine kinase (Kaplan *et al.*, 1994). This pocket lies in the SH2 domain and is known to play a crucial role in stabilizing the kinase in the inactive state (Sicheri, Moarefi, & Kuriyan, 1997).

The overall analysis reveals that the RD-pocket is a spatial motif which is conserved around regulatory phosphorylation sites that bring about conformational changes in proteins. Furthermore, we suggest that incorporation of the spatial motif as a structural feature in kinase phospho-site prediction tools can be used to identify novel allosteric sites regulated by phosphorylation.

CHAPTER 2

RESULTS

2.1 Sequence and Structural Conservation of the RD-Pocket

The canonical RD-pocket is composed of three basic residues - an Arginine immediately preceding the invariant Aspartate (RD-Arg), a basic residue from the beta 9 sheet (DFG+3 Arg/Lys) and a positively charged residue from the alpha C helix (Figure 2.1) (L N Johnson *et al.*, 1996). Although the RD-pocket is preserved to coordinate with the phosphate group, its assembly is not absolutely conserved. Here we have performed a systematic analysis to quantify level of conservation of the RD-pocket from both, sequence and structural perspective.

2.1.1 Sequence Conservation of the RD-Pocket

To compare the RD pocket in different kinase families and to gain insights into the level of sequence conservation, we align the catalytic domain of 10764 representative kinase sequences taken from all seven major Kinase groups. We then measure the conservation of basic amino acids (Arginine, Lysine and Histidine) at positions equivalent to canonical RD pocket in CDK (Figure 2.2). We also look at the distribution of each basic residue at the canonical pocket equivalent positions and the overall conservation of the pocket across the seven major groups.

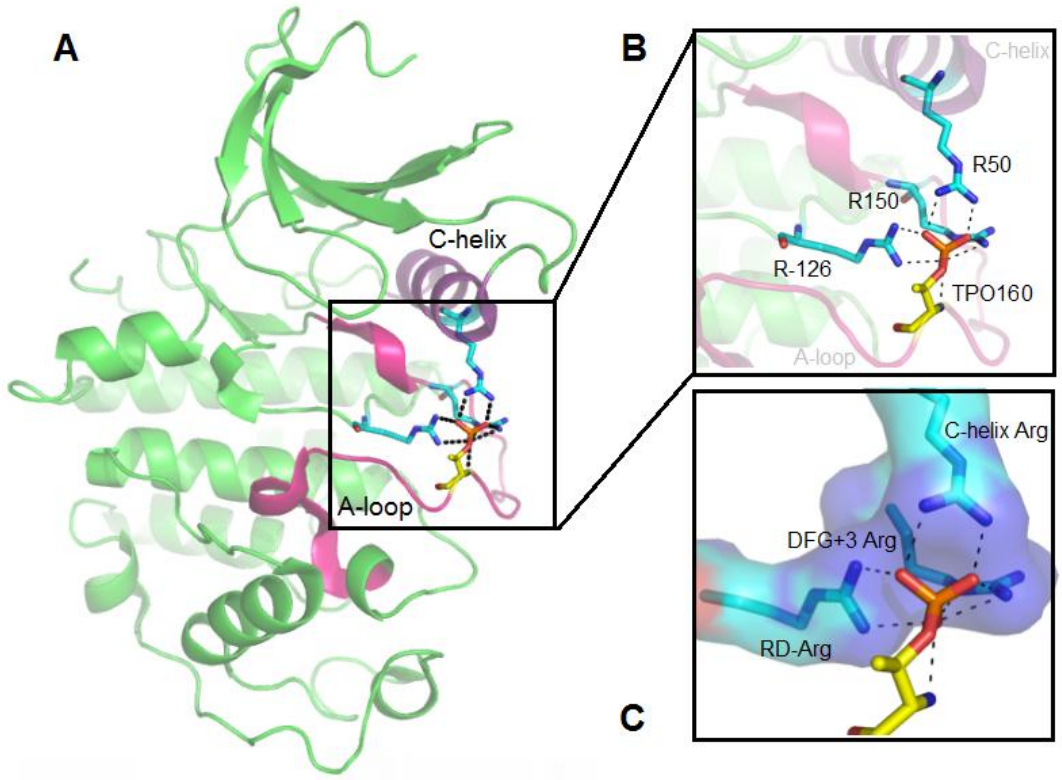


Figure 2.1: Canonical RD pocket from. (A) Overall structure of the catalytic domain of CDK2 (PDB ID: 1QMZ) shown in ribbon diagram. The C-helix is colored deep purple and activation loop is colored pink. (B) Detailed view of the canonical RD-pocket (stick representation, colored blue) and the consensus phosphorylation site in A-loop, TPO160 (stick representation, colored yellow) (C) Surface representation of the canonical RD-pocket.

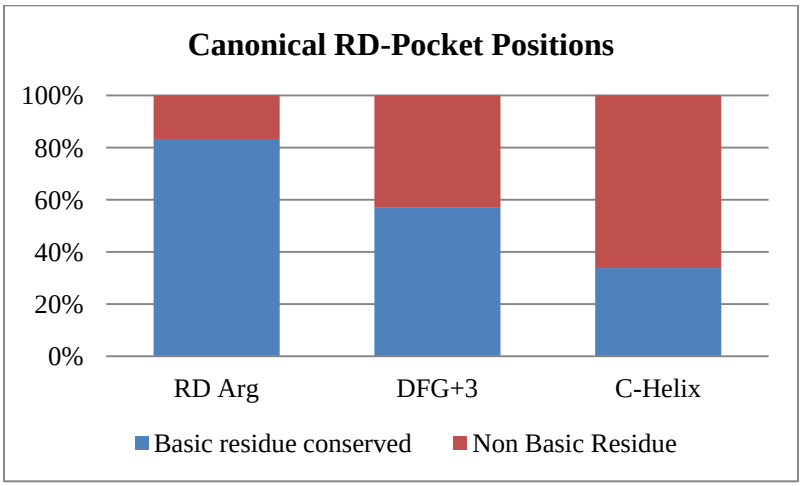


Figure 2.2: Sequence Conservation of the canonical RD-Pocket residues in Kinases

❖ **Conservation of basic residue at RD-Arg position**

Of the three basic residues comprising the RD pocket, the basic residue at the RD-Arg position is most highly conserved (Figure4 A). It is consistently conserved across all seven major kinase families with an overall frequency distribution of 83.32 % (Figure 2.2) in the 10764 representative sequences. The basic residue at the RD-Arg position is predominantly an Arginine (Figure4 B).

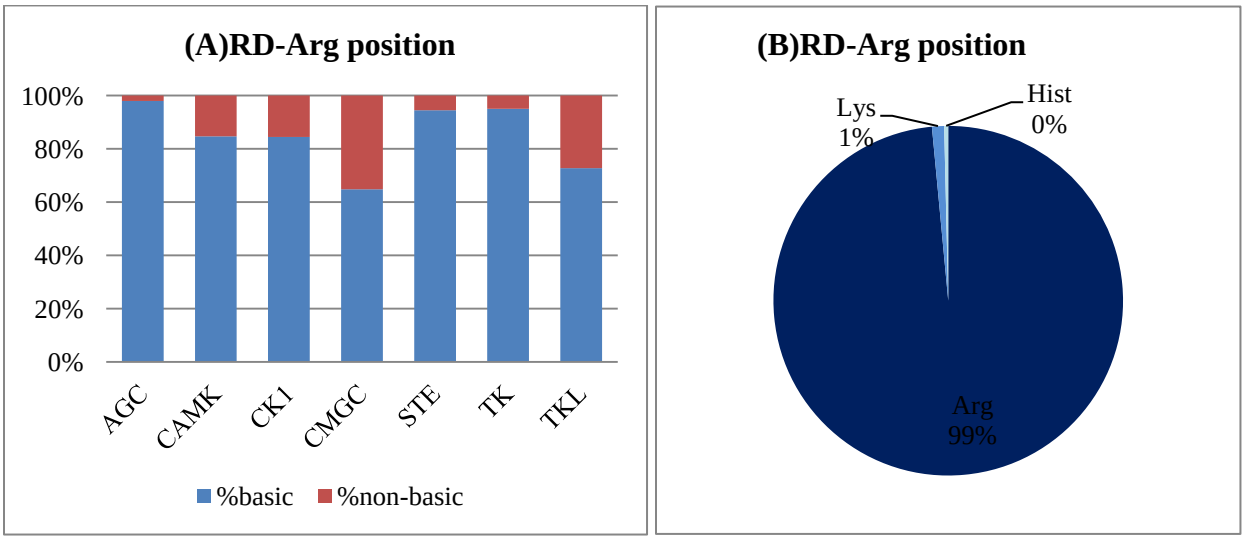


Figure 2.3: Conservation of Basic Residue at RD-Arg position. (A) Conservation of the RD-Arg basic residue across the seven major kinase families. (B) Distribution of the basic residue at the RD-Arg position.

❖ **Conservation of basic residue at DFG+3 position**

The basic residue from the DFG+3 position is the second most highly conserved residue in the RD-pocket. It is a combination of Arginine and a Lysine with an overall frequency distribution of 57.72% (Figure 2.2) (Figure 2.4 B). The DFG+3 Arg/Lys is highly conserved in AGC, CK1, TK & TKL, with frequency distribution of 75.50%, 81.92%, 96.65% and 76.07% respectively (Figure 2.4 A). Conservation of the DFG+3 Arg/Lys is relatively low in CAMK, CMGC and STE with the frequency distribution of 26.58%, 57.87% and 31.03% respectively (Figure 2.4 B).

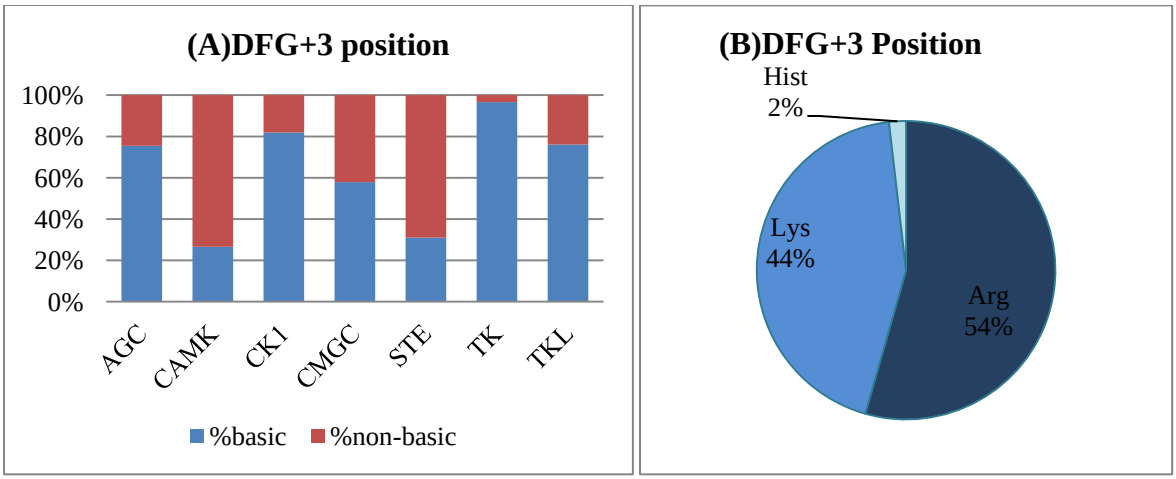


Figure 2.4: Conservation of Basic Residue at DFG+3 position. (A) Conservation of the DFG+3 basic residue across the seven major kinase families. (B) Distribution of the basic residue at the DFG+3 position.

- **Conservation of basic residue at canonical C-helix position:**

The level of conservation of the C-helix basic residue is not as high as that of the RD-Arg or DFG+3Arg/Lys. The overall conservation of the C-helix basic residue in 10764 sequences is 33.82% (Figure 2.2). It can be a combination of Arginine, Lysine or Histidine (Figure 2.5 B). The residue frequency of the specific C-helix basic residue for the kinase families - AGC, CAMK, CK1, CMGC, STE, TK and TKL is 45.34, 32.69, 6.17, 39.57, 20.32, 29.93 and 38.48 percent respectively (Figure 2.5 A).

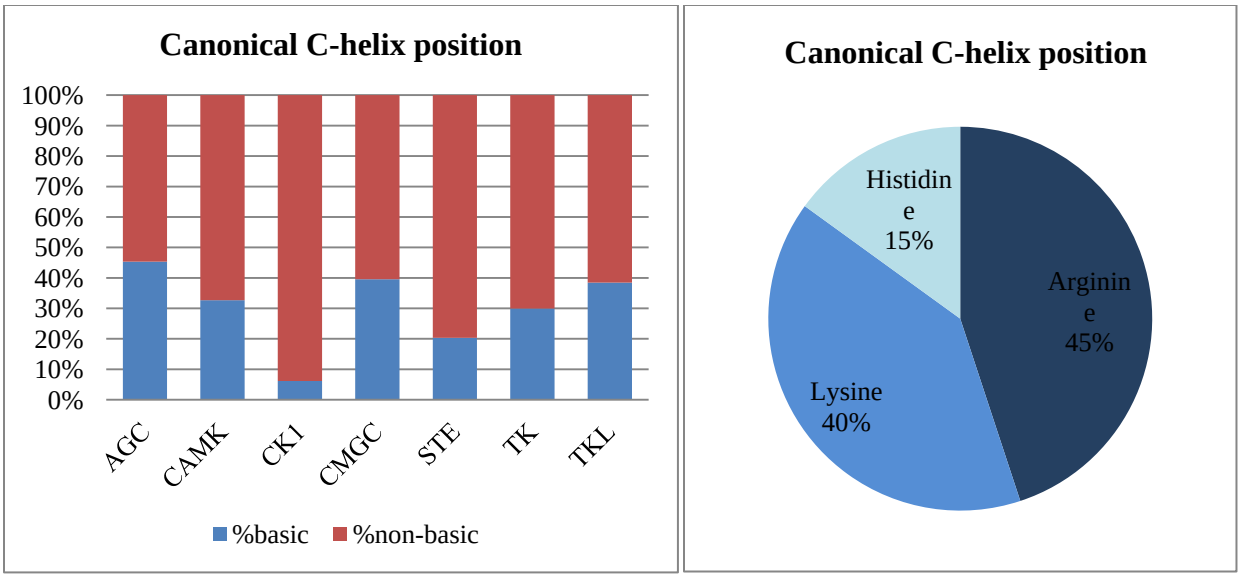


Figure 2.5: Conservation of Basic Residue at canonical C-helix position. (A) Conservation of the basic residue at the canonical C-helix position across the seven major kinase families. (B) Distribution of the basic residue at the canonical C-helix position.

As expected, analysis of the multiple alignments of the kinase sequences shows that the RD Arginine is the most highly conserved component of the RD-pocket (Figure 2.1). Activation loop phosphorylation is a key aspect of regulation in most kinases and it is a well established fact that kinases that require activation loop phosphorylation possess an Arginine immediately preceding the catalytic aspartate (L N Johnson *et al.*, 1996). Kinases that have the RD-Arginine are popularly known as the RD-Kinases (Nolen *et al.*, 2004). Although it is not necessary that all RD-kinases require activation loop phosphorylation, it is highly likely that kinases that lack this RD-Arginine do not require activation loop phosphorylation to acquire activity (L N Johnson *et al.*, 1996); though there are some exceptions to the rule (Kentrup *et al.*, 1996). Furthermore, it is observed that majority of the RD kinases have a basic residue conserved at the DFG+3 position (either an Arginine or a Lysine) and in a few of them the basic residue is

replaced by a non-basic residue. The C-helix Arg/Lys/His is the least conserved residue and is probably the most variable residue in the RD-pocket.

- **Conservation of the overall RD-pocket:**

Next, we look at the conservation of the pocket as a whole. The analysis shows that the canonical RD-pocket is prevalent in members of AGC, CMGC and CAMK family (Figure 2.6). Majority of the members belonging to CK1, STE, TK and TKL groups do not have a canonical pocket. This analysis suggests that more members of the AGC, CMGC and CAMK group are regulated by the canonical phosphorylation mechanism while families within the other groups are not.

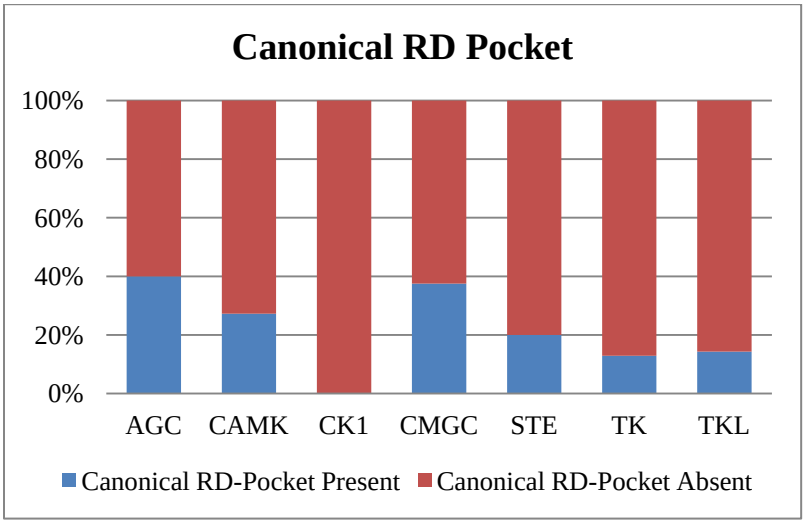


Figure 2.6: Conservation of the canonical RD pocket across seven major kinase families.

2.1.2 Structural Conservation of the RD-Pocket

We now turn to investigate the conservation of the pocket from a structural perspective. Analysis of the structural conservation of the RD pocket is done using SPASM (Kleywegt, 1999) software (see methods). The RD-pocket composed of the RD-Arg, DFG+3 Arg/Lys and C-helix Arg/Lys/His is referred to as “canonical RD

pocket". Using the canonical RD pocket in CDK2 as a query structural motif, a full length library containing 1924 kinase structures is searched. The analysis identifies 74 motifs that are closely related to the query motif based on CDK2 and are within the RMSD of 1 Å (preset cutoff) from the canonical RD-pocket (Table 2.1). The RD-pockets that are formed by basic residues projecting from regions not equivalent to that of the canonical RD-pocket in CDK2 (RD-Arg being conserved) are called "alternative RD-pockets". Pockets that do not contain the RD-Arg are called as "non-RD pockets". Structural comparisons of the identified motifs indicate that several alternative RD-pockets exist and have the same geometry as the query (shown in Figure 2.8 and summarized in Table 2.3). Identification of several alternate RD-pockets suggests that although the RD pocket is not conserved sequentially, it is conserved structurally. Using the canonical RD-pocket as a query, the full length library is searched multiple times, each time with an increased RMSD cutoff (Table 2.1).

The same analysis is repeated with a representative kinase library containing 160 kinase structures (for detailed results see Supplementary Table in Appendices). With 1Å cut-off 28 RD-pockets are identified, out of which 18 are canonical and 10 are alternate RD pockets (Table 2.2). The number of identified RD-pockets increased with lower RMSD cutoff.

Table 2.1: Pockets identified by SPASM in the full length library (1924 structures)

RMSD(Å)	**RD pocket**			**Non- RD Pocket**	**Total Pockets**
	Total	Canonical	Alternate		
1	74	35	39	57	131
1.25	159	56	103	335	494
1.35	209	77	132	537	746
1.4	246	92	154	683	929

Table 2.2: Pockets identified by SPASM in the representative library (160 structures)

RMSD(Å)	RD pocket			Non- RD Pocket	Total Pockets
	Total	Canonical	Alternate		
1	28	18	10	29	57
1.25	56	42	14	125	181
1.35	69	50	19	150	219
1.4	93	69	24	216	309

❖ **Structural Interpretation of Alternative RD Pockets**

Structural comparison of the various RD-pockets reveals that although the RD Arginine is absolutely conserved in almost all these pockets, the location of the second and third basic residue can be variable. Take for example FGFR, JAK2 and IRAK4 kinases, their RD pocket comprises of the RD-Arginine, an Arg/Lys from the DFG+3 position and a basic residue one to two amino acids C-terminus of the phospho-residue in the A-loop (Figure 2.8D). In these families, the basic residue in the C-helix is not conserved. In some families such as PLK and Aurora, the RD-Arg, a basic residue from the C-helix and a basic residue positioned two amino acids N-terminus of the phospho-residue constitutes the RD-pocket (Figure 2.8B). The DFG+3 basic residue of PLK and Aurora is replaced by Thr and Val respectively. In CAMKL & PKR kinases, the RD pocket is composed of the RD-Arg, a Lysine from the C-helix and another Arg from the C-helix (Figure 2.11). In these families the basic residue at the DFG+3 position is replaced by an Asn and a Thr respectively.

In MAPK14, all the canonical RD-pocket residues are present, i.e. the RD-Arg, DFG+3-Arg and the C-helix-Arginine. Interestingly the RD-Arginine does not directly participate in the stabilization of the phosphate moiety, and rather it interacts with the backbone of the phosphorylated threonine (Figure 2.8C). However a basic pocket does

exist, the DFG+3-Arg and two arginines from the C-helix form an alternative RD-Pocket. The structure has 2.1Å resolution and the electron density of the covalently bound phosphate and the phosphorylation site residues are well resolved.

Another point to note is that the RMSD of most of the alternative pockets from the canonical pocket is below 1Å, confirming that the spatial arrangement of all these pockets is very similar. JAK and FGFR have an alternative RD-Pocket, however the overall geometry is slightly different from the canonical RD-pocket in CDK2 as reflected by the increased RMSD (Table3). The slightly altered geometry is presumably because the phosphotyrosine (PTR) is not in the corresponding position of phosphothreonine (TPO) in Ser/Thr kinases; it is two residues N-terminus of the primary phosphorylation site in Ser/Thr kinases (Figure 2.7).

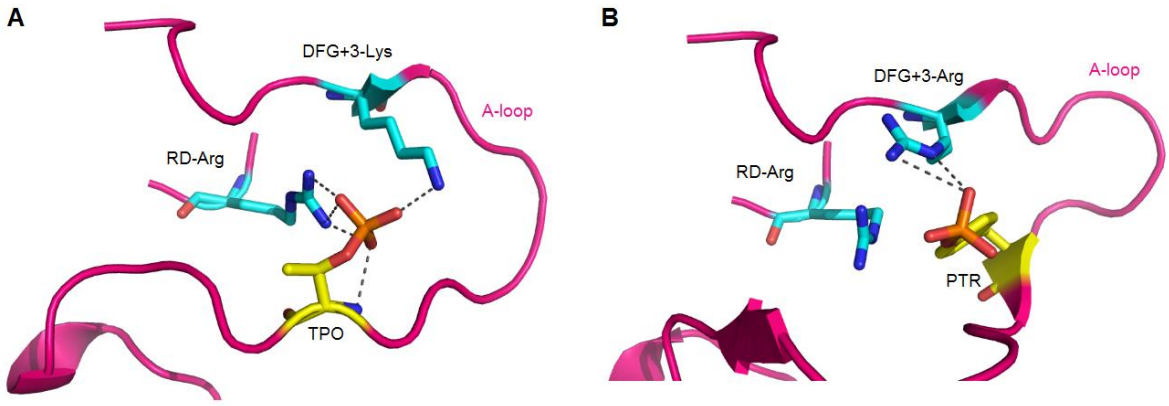


Figure 2.7: Phosphorylation site in Ser/Thr and Tyrosine Kinases. (A) Activation loop of Serine/Threonine kinase (PDB ID: 1ATP). (B) Activation loop of Tyrosine Kinase. Phosphothreonine (TPO) and Phosphotyrosine (PTR) are shown in stick representation and colored yellow.

The above examples show that several alternative RD pockets exist. Although the RD-pocket in general is conserved; the location of the RD pocket is subject to

variability. The pocket is not conserved at sequence level due to the variability in its location. A diverse combination of basic residues, emanating from either the C-helix and/or the A-loop along with the RD Arginine lead to the formation of these distinct alternative RD-pockets, which is more like a spatial motif.

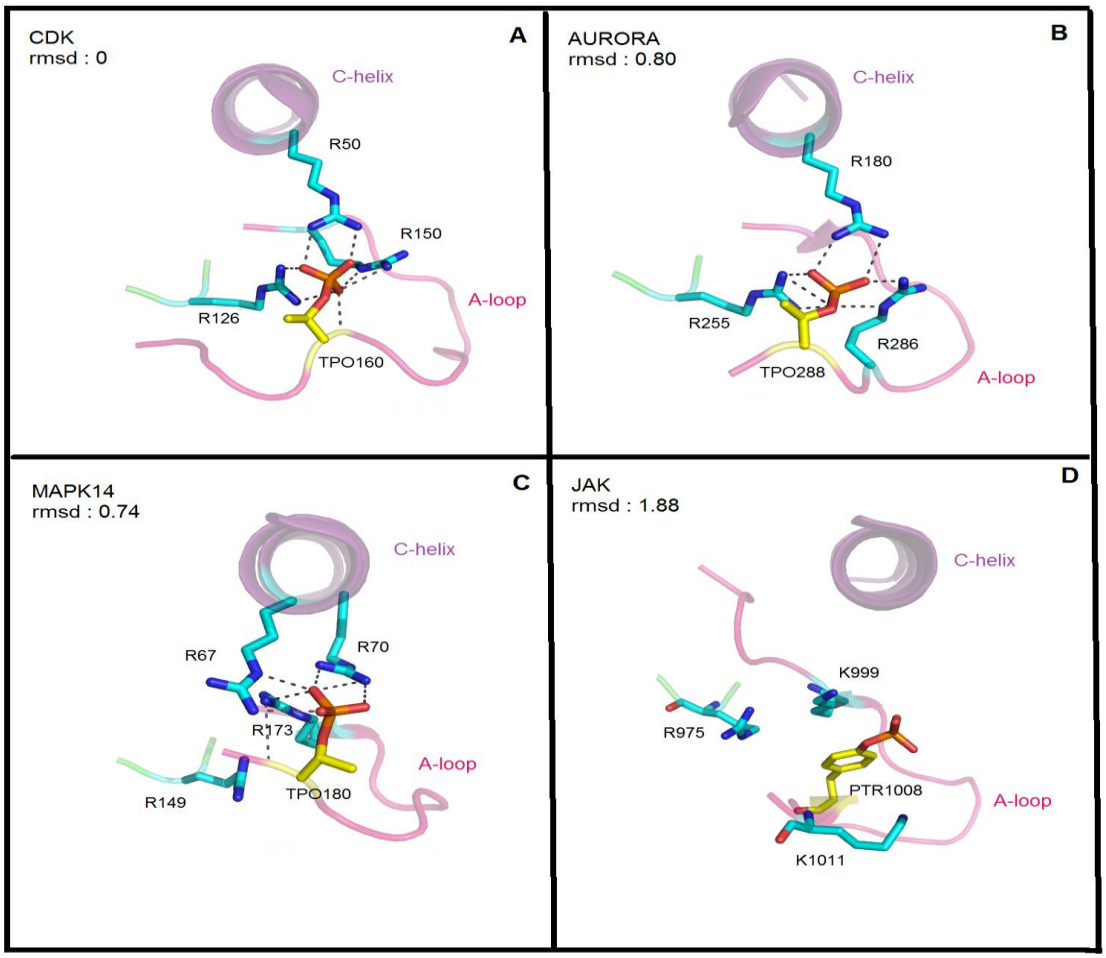


Figure 2.8: Diverse RD-Pockets (A) Close up view of the canonical RD-pocket in CDK2. (PDB ID: 1QMZ) (B) Close up view of the alternate RD-pocket in Aurora A (PDB ID 3HA6) (C) Close up view of the alternate RD-pocket in MAPK14 (PDB ID 3PY3) (D) Close up view of the alternate RD-pocket in JAK (PDB ID 3E63). Coloring scheme is the same as in Figure 2.1.

Table 2.3: Distinct RD-Pockets. Table lists a few alternate RD-pockets identified using canonical RD-pocket as a query. RMSD column tells the RMSD of the identified pocket from the canonical RD pocket in CDK. C-helix column refers to the CDK equivalent residue in C-helix. Other Region residue refers to residues that are not in contact with the primary phosphate in CDK.

Kinase	Group	PDB	RMSD (Å)	Primary Phosphate	RD pocket – basic residues in contact with the primary phosphate			
					RD-Arg	DFG+3	C-helix	Other Region residues
CDK2	CMGC	1QMZ	0	TPO160	Arg126	Arg150	Arg50	NC*
PKA	AGC	1ATP	0.67	TPO197	Arg165	Lys189	His87	NC*
CAMKL	CAMK	2Y94	0.89	TPO172	Arg138	Asn162*	Lys60	R63 - C helix
AuroraA	Other	3HA6	0.80	TPO288	Arg255	Val279*	Arg180	R286 - Aloop , TPO - 2
FGFR	TK	2PVF	1.67	PTR657	Arg625	Arg649	NC*	Lys659, PTR+2
JAK2	TK	3E63	1.88	PTR1008	Arg975	Lys999	NC*	K1011 - Aloop, PTR+3
IRAK4	TKL	2NRY	0.96	TPO345	Arg310	Arg334	NC*	Arg347-Aloop, PTR+2
PLK	Other	3D5W	0.93	TPO196	Arg161	Thr185*	Lys83	K194 - A-loop , TPO-2
PKR	Other	2A19	0.94	TPO446	Arg413	Thr437*	Lys304	Arg307-C helix
MAPK14	CMGC	3PY3	0.74	TPO180	Arg149	Arg173	Arg67	Arg70-C helix
					* Not Contributing to the pocket ~ Arg/Lys replaced			

❖ **Functional Relevance of the Pocket**

The diversity of the RD-pockets observed across these families is a perfect reflection of the diverse regulation mechanisms adopted by different kinase families. While activation loop phosphorylation is a common mechanism of regulation, the specific conformational changes it brings about in a kinase can be family specific. Moreover, the location of the RD pocket elucidates how activation loop phosphorylation induces tertiary and quaternary structural changes. This is exemplified by the diverse alternative RD-pockets in FGFR, Aurora A and PKR, described below.

Role of the Alternative RD-Pocket in FGFR: In FGFR the RD-pocket involves the RD-Arg625, DFG+3 Arg649 and Lys659 from the A-loop (PTR+2 position). The pocket plays a crucial role in releasing the autoinhibitory conformation of the A-loop upon phosphorylation. Unlike CDK, in the unphosphorylated state of FGFR, the C-helix obtains a closed-like conformation and the Lys-Glu salt bridge is intact. The Lys-Glu salt bridge is crucial for ATP-binding and is always intact in the active state confirmation (Hanks & Hunter, 1995). However, the activation loop is highly mobile and plays an autinhibitory role as its C-terminal end blocks the substrate binding site, see Figure 2.9 A (Mohammadi, Schlessinger, & Hubbard, 1996). Upon phosphorylation the A-loop is released from the autoinhibitory conformation (Figure 2.9 B). The phosphorylated active state conformation is stabilized primarily by hydrogen bonds between phosphate group of PTR657 and the side chains of two basic from the A-loop: R649 and K659. R648 and K659 form the alternative RD pocket together with the RD-Arginine (Figure 2.9 C) (H. Chen *et al.*, 2007). Here the interaction with the K659 is important as it helps demobilize

the activation loop and stabilize the active state conformation of A-loop (H. Chen *et al.*, 2007).

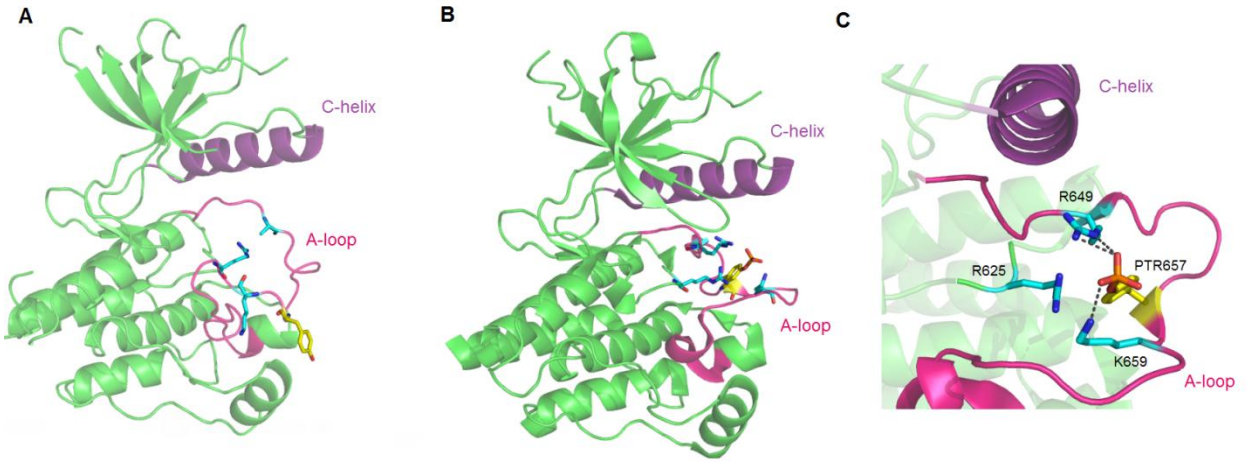


Figure 2.9: Alternate RD Pocket in FGFR. (A) Overall structure of inactive fgfr in ribbon diagram (PDB ID 1FGK). (B) Overall structure of active fgfr in ribbon diagram (PDB ID 2PVF) (C) Close up view of the alternate RD-pocket in FGFR. Coloring scheme is the same as in Figure 2.1.

Role of the Alternative Pocket in Aurora A: In Aurora kinases the RD-pocket plays a key role in the activation of the enzyme. It is composed of RD-Arg255, C-helix Arg180 and Arg286 from the A-loop (TPO-2) position. Acquisition of activity in Aurora A is a two step process that requires A-loop phosphorylation and TPX2 binding (Eyers, Erikson, Chen, & Maller, 2003; Kufer *et al.*, 2002; Littlepage *et al.*, 2002; Walter, Seghezzi, Korver, Sheung, & Lees, 2000). Phosphorylated A-loop in Aurora A is in the inactive conformation in the absence of TPX2 as the phosphothreonine (TPO288) does not bind to alpha-C-helix Arg180 (Figure 2.10 A) (Bayliss, Sardon, Vernos, & Conti, 2003). This is because the alternate RD-pocket is not formed in the unbound TPX2 state as the side chain of R286 is solvent exposed and that of R180 is disoriented. Due to the absence of the pocket, TPO288 is solvent exposed rather than pointing towards RD-Arg255. The activation loop in turn blocks with the substrate binding side rather than providing a

substrate binding platform. TPX2 binding reorients the side chains of Arg180 and Arg286 (Bayliss *et al.*, 2003) such that an alternative phosphate recognizing RD-pocket is formed (Figure 2.10 B). The phosphothreonine now establishes crucial contacts with all three Arginines of the alternative RD pocket required to achieve the active state confirmation.

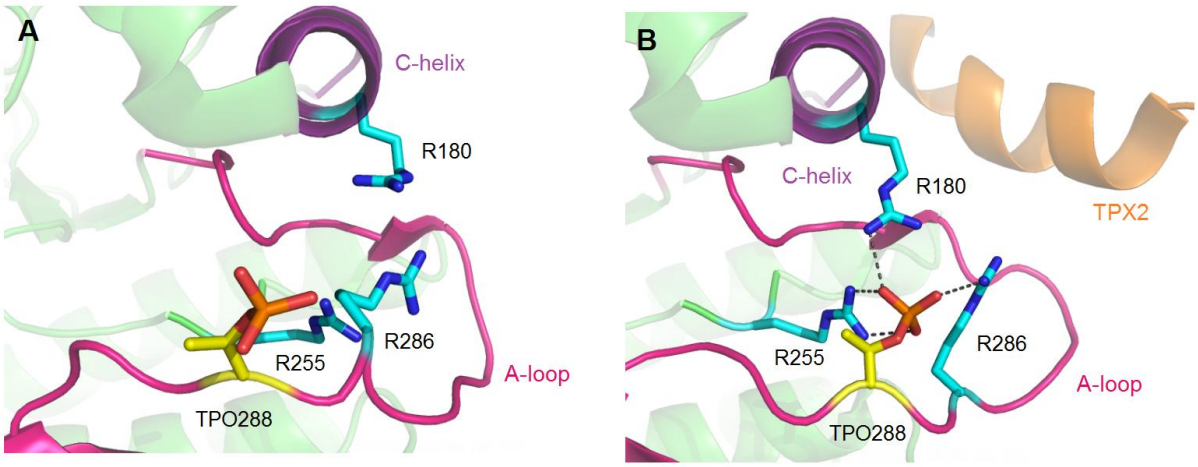


Figure 2.10: Alternate RD-pocket in Aurora. (A) Close up view of the phosphorylated activation loop in TPX2-unbound state (PDB ID: 1OL7) (B) Close up view of the alternate RD-pocket in phosphorylated TPX2-bound state of Aurora. TPX2 is colored orange and the rest of the coloring scheme is same as figure1.

Role of the Alternative Pocket in PKR: In RNA dependent protein kinase PKR the pocket is composed of RD-Arg413, Lys304 and Arg307 projecting from the C-helix (Figure 2.11). This interaction between the pocket and the phosphate moiety of TPO437 stabilize the active state conformation of the enzyme. The alternative pocket in PKR is different from the canonical RD-pocket in that it does not involve the canonical basic residue from the activation loop (DFG+3). One possibility for the alternate pocket is that the unique conformation of the P+1 loop in the active state. Unlike most Ser/Thr kinases

where the P+1 loop adopts an inward position relative to the catalytic cleft, in PKR the P+1 loop adopts an extreme outward orientation that is similar to Tyrosine Kinases (Dar, Dever, & Sicheri, 2005). It has been suggested that this distortion in the P+1 loop allows the dual specificity of the kinase (Dar *et al.*, 2005). The alternate RD-pocket allows the enzyme to obtain the non-canonical P+1 conformation.

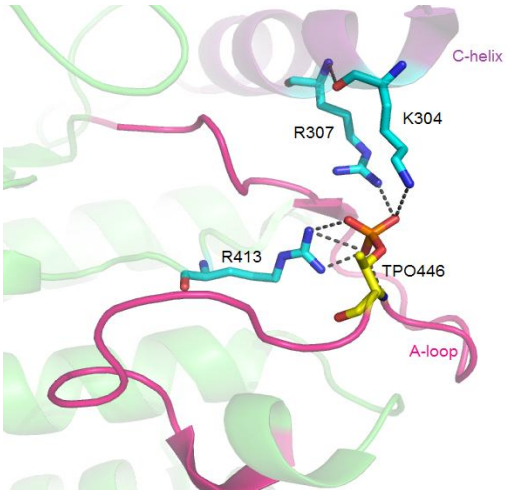


Figure 2.11: Alternate RD-pocket in PKR. Color scheme is same as Figure 2.1.

2.2 Non- RD Pockets

A close inspection of the motif hits identified upon increasing the RMSD cut-off reveals that several RD-pocket like pockets are present in other regions of the kinase that does not encompass the activation loop consensus phosphorylation site or the RD-Arg. Such basic pockets similar to RD-pocket are termed as “non-RD pockets”. With the full length library and RMSD cutoff equal to 1Å, total number of pockets identified by SPASM (Kleywegt, 1999) is 131. The majority of hits obtained are RD-pockets: 38 canonical RD-pockets and 43-alternate RD pockets. However 57 non-RD pockets are also identified. As the RMSD cut off for motif finding is increased to 1.25, the number of

non-RD pockets identified by SPASM (Kleywegt, 1999) increases by 278, as shown inTable 2.1.

We next wanted to determine the functional significance of these pockets. A systematic study of these pockets reveals that non-RD-pockets are present around regulatory phosphorylation sites other than that in the activation-loop. Furthermore, our analysis suggests that the pockets can be used to identify potential allosteric sites that are regulated by phosphorylation. Some of the most significant hits obtained by SPASM (Kleywegt, 1999) are described here.

2.2.1 Non-RD Pocket in Src Kinase

A non-RD pocket is identified in the SH2 domain of Src-Kinase. The pocket is composed of Arg155, Arg175 and His201 (Figure 2.12 B) and its RMSD from the canonical RD-pocket is 0.83Å. The pocket encompasses a phosphotyrosine PTR527 emanating from the C-terminal tail of the kinase domain. The non-RD pocket is absolutely conserved in all vertebrates and invertebrates (Figure 2.12 C). Tyrosine 527 (Y527) is a well known phosphorylation site of Src kinases and its phosphorylation negatively regulates the activity of the enzyme (Roskoski, 2004; Sicheri *et al.*, 1997). The basic pocket acts as a phosphate binding site and the interactions between PTR527 and the pocket are crucial for holding the enzyme in an inactive state (Roskoski, 2004; Sicheri *et al.*, 1997).

Src kinases belong to non-receptor tyrosine kinase family and are the first described member of this family. They contain a tyrosine kinase domain (TKD), a short carboxy-terminal tail and catalytically inactive SH2 and SH3 domain (Figure 2.12 A). The regulation mechanism of the Src kinase catalytic activity has been studied

comprehensively (Superti-Furga & Courtneidge, 1995). Two phosphorylation events control the activity of the kinase domain – autophosphorylation of the activation loop at Y416 & phosphorylation of the C-terminal tail segment at Y527 (Roskoski, 2004; Zheng, Resnick, & Shalloway, 2000). While autophosphorylation of the activation loop up regulates the activity, phosphorylation of the c-terminal segment by Csk down regulates the activity (Roskoski, 2004; Zheng *et al.*, 2000). The phosphorylated Y527 binds to the SH2 domain by establishing intramolecular interactions with the non-RD basic pocket in the SH2 domain (Sicheri *et al.*, 1997). In the catalytically active form, the TKD interacts only with the SH3-domain (Cowan-Jacob *et al.*, 2005). Phosphorylation of the C-terminal tail induces large conformational change in all three domains and results in substantial movements of the SH2-SH3 domain. The non-RD pocket plays a crucial role as it is contributes immensely to the high affinity intramolecular interactions between the phosphorylated c-terminal tail and the SH2 domain. These intramolecular interactions stabilize the dormant form of the enzyme (Wenqing & Harrison, 1997; Williams *et al.*, 1997).

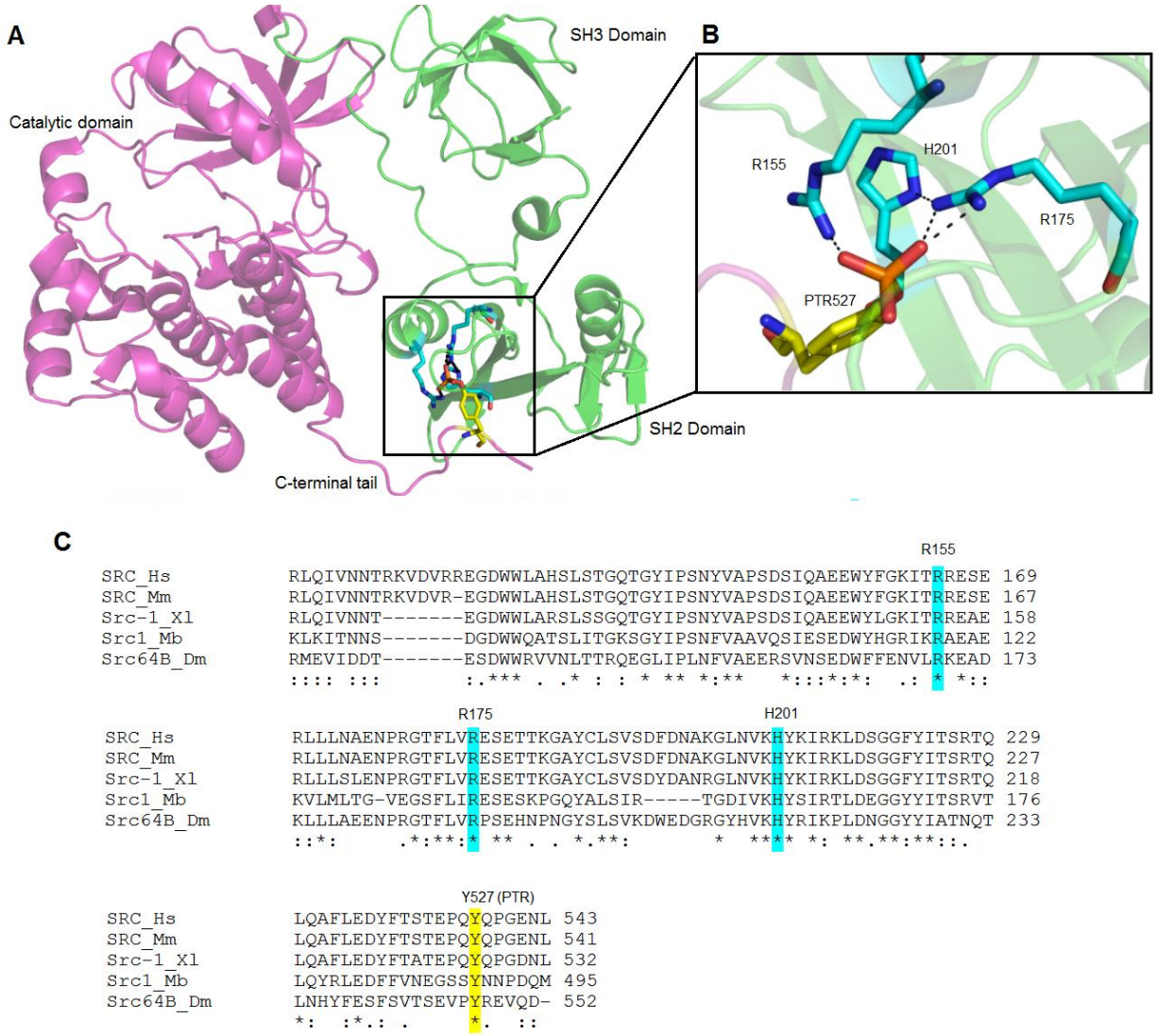


Figure 2.12: Non-RD-pocket in SRC. (A) Overall structure of an inactive Src kinase (PDB ID 2H8H) is shown in ribbon diagram. The inactive conformation is stabilized by intramolecular interactions between the catalytic domain and SH2-SH3 domain. (B) A detailed view of the interactions between the non-RD pocket in SH2 domain (stick representation, colored cyan) and phosphotyrosine (stick representation, colored yellow) from the C-terminal tail. (C) ClustalW sequence alignment of non-RD pocket containing region in SRC of *human* (Hs), *mouse* (Mm), *Xenopus* (Xl), *M. brevicollis* (Mb) and *Drosophila* (Dm). Basic residues forming the non-RD pocket are highlighted in blue and the phosphorylation site is highlighted in yellow.

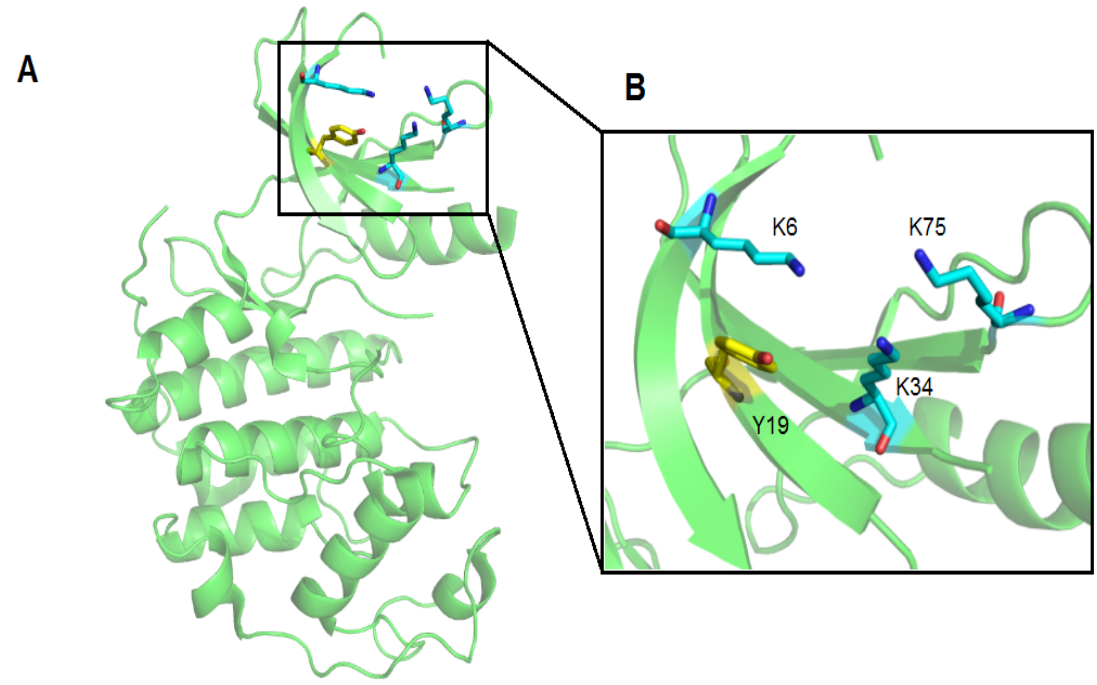
2.2.2 Non- RD Pocket in CDK2

Another interesting non-RD pocket was identified in CDK2 family. The pocket is composed of three Lysines ~ K6, K34 & K75 (Figure 2.13 B) and its RMSD from the canonical RD-pocket is 1.22 Å. The pocket is located on the N-terminal region (Figure 2.13 A), on beta sheets and is conserved in all vertebrates (Figure 2.13 C). In some lower organisms such as *Drosophila*, K75 in the basic pocket is replaced by N75. The predicted phosphosite Tyr19 is enclosed by this basic pocket. Although no crystal structure has been solved with a phosphorylated Tyr-19, several mass spectrometry studies provide evidence that Tyr19 gets phosphorylated (Oppermann *et al.*, 2009; Rikova *et al.*, 2007; Wissing *et al.*, 2007). In fact, cell-cycle regulated dephosphorylation of Tyr-19 has been shown to be required for Spindle Pole Body Segregation (Lim, Goh, & Surana, 1996). Here we propose that the identified non-RD pocket is the phosphate binding site that packs together to stabilize the phosphate moiety once Tyr19 is phosphorylated. This rearrangement of the pocket upon phosphorylation likely leads to quaternary conformational changes associated with SPB segregation defect.

Cyclin Dependent Kinases (CDKs) complexed with various cyclins are known to play a major role in controlling the cell cycle in eukaryotic organisms (Morgan, 1997; Murray, 2004). Duplication of the centrosome followed by their segregation is one of the many cell cycle events regulated by CDKS (Haase, Winey, & Reed, 2001; Lacey, Jackson, & Stearns, 1999). Animal cell centrosomes are functionally equivalent to yeast spindle pole bodies (SPB). Mutational analysis of CDK2-Tyr19 in *Saccharomyces Cervisiae* revealed the importance of its dephosphorylation in SPB separation. Tyr19 to glutamate mutation mimicked phosphorylation induced inhibition and resulted in SPB

segregation defect (Lim *et al.*, 1996). This phospho-mimetic mutation is thought to inhibit complex formation of CDK with Cyclin-B homologs. (Keaton *et al.*, 2007), which is in turn diminishes SPB segregation as Cyclin-B-homolgs are involved in SPB segregation process (Fitch *et al.*, 1992; Richardson, Lew, Henze, Sugimoto, & Reed, 1992). The identified basic pocket is not very well formed (RMSD from canonical RD-pocket is 1.52 Å), which is not surprising given that Tyr-19 is not phosphorylated. Once Tyr19 is phosphorylated, it is highly likely that Lys75 that lies on the loop joining beta4 & beta5 sheets undergoes a conformational change. The induced conformational change is likely to bring Lys75 in close proximity of the phosphate moiety in order to establish electrostatic interactions with it and stabilize its negative charge. This conformational change could bring about other quaternary changes in the kinase domain that inhibit complex formation with Cyclin – B homologs; in turn leading to SPB segregation defect.

We also attempted to see if this phosphorylation site is detected by phospho-site prediction algorithms. While NETPHOS 2.0 (N Blom *et al.*, 1999) and PPSP (Xue *et al.*, 2006) do not predict Y19 as a phosphorylation site , DISPHOS (Iakoucheva *et al.*, 2004) does predict Y19 as a phosphorylation site.



C

```
                      K6                  Y19                 K34
CDK2_Hs       ----MENFQKVEKIGEGTYGVVYKARNKLTGEVVALKKIRLDTETEGVPSTAIREISLLK 56
CDK2_Mm       ----MENFQKVEKIGEGTYGVVYKAKNKLTGEVVALKKIRLDTETEGVPSTAIREISLLK 56
CDK2_Xl       ----MENFQKVEKIGEGTYGVVYKARNRETGEIVALKKIRLDTETEGVPSTAIREISLLK 56
CDK2_Sp       ----MNNFEKIEKIGEGTYGVVYKARDLKSGKTVALKKIRLDTESEGVPSTAIREIALLK 56
cdc2c_Dm      MTTILDNFQRAEKIGEGTYGIVYKARSNSTGQDVALKKIRLEGETEGVPSTAIREISLLK 60
                  ::**:: *********:*:**:.  :*: ******:**: *:***********:***

                                 K75
CDK2_Hs       ELNHPNIVKLLDVIHTENKLYLVFEFLHQDLKKFMDASALTGIPLPLIKSYLFQLLQGLA 116
CDK2_Mm       ELNHPNIVKLLDVIHTENKLYLVFEFLHQDLKKFMDASALTGIPLPLIKSYLFQLLQGLA 116
CDK2_Xl       ELNHPNIVKLLDVIHTENKLYLVFEFLNQDLKKFMDRSNISGISLALVKSYLFQLLQGLA 116
CDK2_Sp       ELDHKNIVKLHDVVHSDKKLYLVFEFMNQDLKKYMDVAPPSGLPPGLVKSYLQQLLQGIA 116
cdc2c_Dm      NLKHPNVVQLFDVVISGNNLYMIFEYLNMDLKKLMDKKK-DVFTPQLIKSYMHQILDAVG 119
              :*.* *:*:* **: : ::**::**::: **** **      :.  *:***: *:*:.:.
```

Figure 2.13: Non-RD-pocket in CDK2. (A)Overall structure of CDK2 (PDB ID 1G5S) is shown in ribbon diagram (B) Detailed view of the non-RD pocket (stick representation, colored cyan) and the detected phospho-site Tyr19 (stick representation, colored yellow) (C) ClustalW sequence alignment of CDK2 kinase domain containing the non-RD pocket of *human* (Hs), *mouse* (Mm), *Xenopus* (Xl), *Sea Urchin* (Sp) and *Drosophila* (Dm). Basic residues forming the non-RD pocket are highlighted in blue and the phosphorylation site is highlighted in yellow.

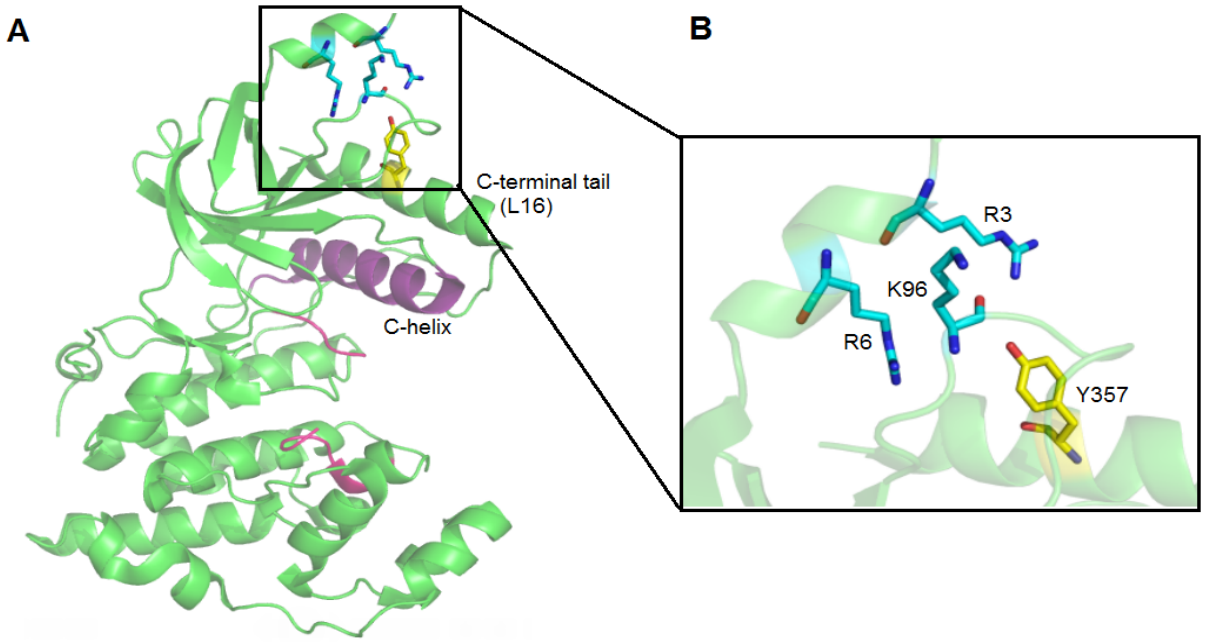
2.2.3 Non- RD Pocket in MAPK8

The next pocket composed of Arg3, Arg6 and Lys96 is identified at the N-terminus end of MAPK8 (JNK1) and it envelopes Tyr357 located at the C-terminal tail (L16) of the enzyme (Figure 2.14 A, B). The RMSD of the identified non-RD pocket from the canonical RD pocket is 1.43Å. The non-RD-pocket is conserved in vertebrates and there are some interesting variations in the lower organisms (Figure 2.14 C). For example in *B. malayi* and *C. elegans*, Arg3 and Arg6 are not conserved and curiously Tyr357 is also replaced by Phe357. Phosphoproteomic studies provide evidence of phosphorylation at Tyr357 (Yu *et al.*, 2007), although a crystal structure with phosphorylated Tyr357 is unavailable. Also, the functional significance of this phosphorylation site has not been explored so far. Here we propose that Tyr357 is a key regulatory site. Interactions between PTR357 and the identified non-RD pocket can promote lobe closure and lock the kinase in an ATP-binding competent conformation.

MAPKs are present in all eukaryotic organisms and multiple MAPK modules are used for transduction of signals originating from the receptors at the cell surface (Lewis, Shapiro, & Ahn, 1998; Robinson & Cobb, 1997). In MAPKs the N-terminal domain is composed of the N-lobe and the C-terminal tail (L16). L16 lies on the surface of the N-terminal domain and has been shown to be critically important for the function of MAPK6, MAPK7 and MAPK15 (Abe, Kuo, Hershenson, & Rosner, 1999; Aberg *et al.*, 2009; Buschbeck & Ullrich, 2005). The C-terminal tail of other MAPK subfamily members has also been suggested to be important for their function (Li, Liu, & Zhang, 2011). However, the role of L16 is different for different MAPK subfamilies. For instance, the C-tail In MAPK7 plays an inhibitory role (Buschbeck & Ullrich, 2005), while in

MAPK15 it regulates cellular localization (Abe *et al.*, 1999). The role of C-tail in MAPK8 is not very well understood. The presence of a non-RD pocket around the Tyr357 suggests that the pocket is a phosphate recognizing pocket. As observed in Figure 2.14 B, in the unphosphorylated form the pocket is not very well formed, the Lys96 side chain is pointing away from Tyr357. Phosphorylation of Tyr357 is likely to induce conformational changes so that the side chains of the basic residues emanating from flexible loop regions adjust their spatial orientation in order to establish hydrogen bonds with the phosphate moiety. As the L-16 lies almost parallel to the C-helix, the phosphorylation induced conformation change can have a direct impact on C-helix orientation. This in turn, may facilitate activation by promoting lobe closure.

MAPK8 signaling pathway is known to be regulated by JNK-interacting-protein-1(JIP1) (Dickens *et al.*, 1997). Overexpression of JIP1 inhibits the activity of MAPK8 and structures of MAPK8-JIP1 complex have revealed an allosteric inhibition mechanism (Dickens *et al.*, 1997; Heo *et al.*, 2004). Binding of JIP1 induces a hinge motion between the N-terminal and C-terminal domains. This in turn distorts the ATP binding cleft, which reduces the affinity of the enzyme for ATP (Heo *et al.*, 2004). As L16 lies is located parallel on top of C-helix, phosphorylation of C-tail could lock the kinase in a high ATP-binding competent conformation and its dephosphorylation can make the enzyme susceptible to JIP1 binding. Thus, Tyr357 could be an important regulatory site of MAPK8. Other phospho-site prediction algorithms like NETPHOS 2.0 (N Blom *et al.*, 1999), DISPHOS (Iakoucheva *et al.*, 2004) and PPSP (Xue *et al.*, 2006) also predict Tyr357 as a phosphorylation site.



C

```
                                                    R3  R6
JNK1_Hs         ---------------------------------MSRSKRDNNFYSVEIGDSTFTVLKRYQ 27
JNK1_Mm         ---------------------------------MSRSKRDNNFYSVEIGDSTFTVLKRYQ 27
JNK1_Xl         ---------------------------------MSRSKRDSNFSVFEIGDSTFTVLKRYQ 27
JNK1_Dr         ---------------------------------MNRNKREKEYYSIDVGDSTFTVLKRYQ 27
JNK1_Bm         YETEAILLDSCASVLQPS------YLEGVYPASMASTSTASEFYNVDINDTRLCVLKRYQ 72
JNK1_Ce         RDDDAVSLDSCQSVYSPVRHHINSGTGGGILMEPSSIHVPENYYSVTIGEAQMVVLKRYQ 120
                                                    .::  . :.:: : ******
                      K96
JNK1_Hs         LLNVFTPQKSLEEFQDVYIVMELMDANLCQVIQMELDHERMSYLLYQMLCGIKHLHSAGI 147
JNK1_Mm         LLNVFTPQKSLEEFQDVYIVMELMDANLCQVIQMELDHERMSYLLYQMLCGIKHLHSAGI 147
JNK1_Xl         LLNVFTPQKSLEEFQDLYIVMELMDANLCQVIQMELDHERMSYLLYQMLCGIKHLHSAGI 147
JNK1_Dr         LLNVFTPQKTLEEFQDVYLVMELMDANLCQVIQMELDHERLSYLLYQMLCGIKHLHAAGI 147
JNK1_Bm         LLNAFTPQKTLDEFSDLYIVMELMDANLCQVIQMDLDHERMSYLLYQMLCGIRHLHAAGI 192
JNK1_Ce         ILNCFTPQKKLDEFNDLYIVMELMDANLCQVIQMDLDHERLSYLLYQMLCGIRHLHSAGI 240
                :** *****.*:**.*:*:***************:*****:***********:***:***

                  Y357
JNK1_Hs         ELIYKEVMDLEERTKN-GVIRGQPSPLGAAVINGSQHPSSSSSVNDVSSMSTDPTLASDT 412
JNK1_Mm         ELIYKEVMDLEERTKN-GVIRGQPSPLAQVQQ---------------------------- 384
JNK1_Xl         ELIYKEVLDWEERAKN-GVIRGQPAPLGAAVTDGSQAHTSSSSG-DASSMSTDPTLPSDT 411
JNK1_Dr         ELIYKEVLEWEERTKN-GVIRGQPASLAQVQQ---------------------------- 384
JNK1_Bm         ELIFKEIMQYEQTHDEYGVKKSLNCTSGAVPSSSS------------SFQNNEHGDCSTT 540
JNK1_Ce         EHIFRELTDYARTHDIYS------------------------------------------ 463
                * *::*: :  .  .  .
```

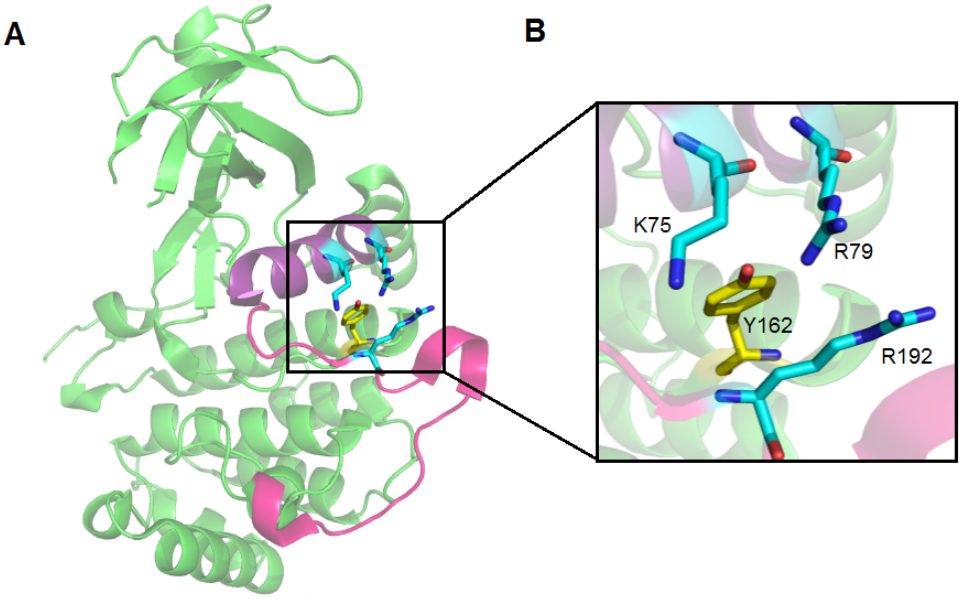
Figure 2.14: Non-RD-pocket in JNK1. (A) Overall structure of MAPK8 (PDB ID 2XRW) is shown in ribbon diagram. C-helix is colored deep purple and A-loop is colored hot-pink. (B) Detailed view of the non-RD pocket (stick representation, colored cyan) and the detected phospho-site Tyr357 (stick representation, colored yellow) (C) ClustalW sequence alignment of non-RD pocket containing region in JNK1 of *human* (Hs), *mouse* (Mm), *Xenopus* (Xl), *Zebra fish* (Dr), *B. malayi* (Bm) and *C. elegans* (Ce). Basic residues forming the non-RD pocket are highlighted in blue and the phosphorylation site is highlighted in yellow.

2.2.4 Non- RD Pocket in VRK2

A non-RD pocket composed of three arginines is found in VRK2, its RMSD from the canonical RD-pocket is 1.47Å. The pocket surrounds Tyr162 that lies right before the catalytic loop. Two arginines from the C-helix and an Arginine from the DFG+4 position form this pocket (Figure 2.15 A, B). The location of the pocket is very close to that of the canonical RD-pocket but it does not involve the RD – Arginine or the consensus primary phosphorylation site. Tyr162 is not known to be phosphorylated. Here, we hypothesize that Tyr162 is a putative phosphorylation site and the impact of its phosphorylation would be comparable to A-loop phosphorylation in Vrk2.

Vaccinia Related Kinase 2 is one among the few kinases that have shown to maintain activity even after loss of key functional residues (Blanco, Klimcakova, Vega, & Lazo, 2006; Nichols, Wiebe, & Traktman, 2006). Structures of such kinases often reveal compensatory changes that help retain the catalytic activity, like in the case of Titin & Wnk (Mayans *et al.*, 1998; Min, Lee, Cobb, & Goldsmith, 2004). However, very little is known about the regulation mechanism of VRK2, and only one crystal structure of Vrk2 has been solved till date (Scheeff, Eswaran, Bunkoczi, Knapp, & Manning, 2009). Proteomic studies indicate that Vrk2 can be regulated by allosteric protein-protein interaction (Sanz-García, López-Sánchez, & Lazo, 2008). Our analysis suggests that Tyr162 can be a key regulatory phosphorylation site of Vrk2. Vrk2 has an “HGD” and “DYG” motif instead of the conventional “HRD” and “DFG” motifs, because of which it can neither form a conventional RD-pocket or be phosphorylated at the primary consensus phosphorylation site. Tyr162 and the observed non-RD pocket formed by R75, R79 and R192 can compensate for the loss of the activation loop phospho-site

and the canonical RD-pocket. So, phosphorylation of Tyr162 could play a regulatory role analogous to that of activation loop phosphorylation. Our results are consistent with other phospho-site prediction algorithms like NETPHOS 2.0 (N Blom *et al.*, 1999), DISPHOS (Iakoucheva *et al.*, 2004) and PPSP (Xue *et al.*, 2006) that also predict Tyr162 as a potential phosphorylation site. Our analysis and the high score for Tyr162 in the phospho-site prediction methods suggest that Tyr 162 is a true phosphorylation site.



C

```
                          K75  R79
VRK2_Hs         VEYQENGPLFSELKFYQRVAKKDCIKKWIERKQLDYLGIPLFYGSGLTEFKGRSYRFMVM 93
VRK2_Mm         LEYQENGPLFSELKFYQRAAKRECIQKWIQQRKLDYLGIPVFYGFGLTDFKGRSYRFMVM 93
VRK2_Xl         VEYHQNGPLFCELKFYQRAAKPDNITKWKAHKKLDYLGIPKYWGSGETTFNSQSYRFMVM 120
VRK2_Dr         VEYHENGPLFSELKFYQRAAKPETMSKWMKSKQLGFLGIPTYWGSGLTESNGTRYRFMVM 120
                :**::*****.*******.** : : **   ::*.:**** ::* * *  :.  ******

                                                          Y162
VRK2_Hs         ERLGIDLQKISGQN-GTFKKSTVLQLGIRMLDVLEYIHENEYVHGDIKAANLLLGYKNPD 152
VRK2_Mm         ERLGIDLQKLLDQN-GGFKKLTVLQLGIRMLDVLEYIHENEYVHGDIKAANLLLDFTNPD 152
VRK2_Xl         DRLGVDLQTILNKNNGKFSMQRVMQLGICMLDVLEFIHENEYVHCDIKAANILLSYTNQT 180
VRK2_Dr         DRLGTDLQKVLIDNGGQLRKTSVLQLGVLMLDVLEYIHDNEYVHADIKAANLLLGYRDPN 180
                :*** ***.:  .* * :    *:***: ******:**:*** ******:**.: :

                     R192
VRK2_Hs         QVYLADYGLSYRYCPNGNHKQYQENPRKGHNGTIEFTSLDAHKGVALSRRSDVEILGYCM 212
VRK2_Mm         RVYLADYGLSYRYCPNGNHKQYQEDPRKGHNGTIEFTSLDAHKGVAPSRRSDVEILGYCM 212
VRK2_Xl         EVYLADYGLSYRYCPNGNHKEYKENPRKGHNGTIEFTSLDAHKGVAPSRRGDLEILAYCM 240
VRK2_Dr         KVYLADYGLSYRYCPNGEHKEYKENPKKGHNGTIEYTSIDAHKGVAASRRGDLEVLGYCL 240
                .**********:*****:**:*:*:*:*:********:**:******* ***.*:*:*.**:
```

Figure 2.15: Non RD-pocket in VRK2. (A)Overall structure of VRK2 (PDB ID 2V62) is shown in ribbon diagram. The C-helix is colored deep purple and A-loop is colored hot pink (B) Detailed view of the non-RD pocket (stick representation, colored cyan) and the potential phospho-site Tyr162 (stick representation, colored cyan) (C) ClustalW sequence alignment of VRK2 kinase domain containing the non-RD pocket of *human* (Hs), *mouse* (Mm), *Xenopus* (Xl) and *Zebra fish* (Dr). Basic residues forming the non-RD pocket are highlighted in blue. Putative phosphorylation site is highlighted in yellow.

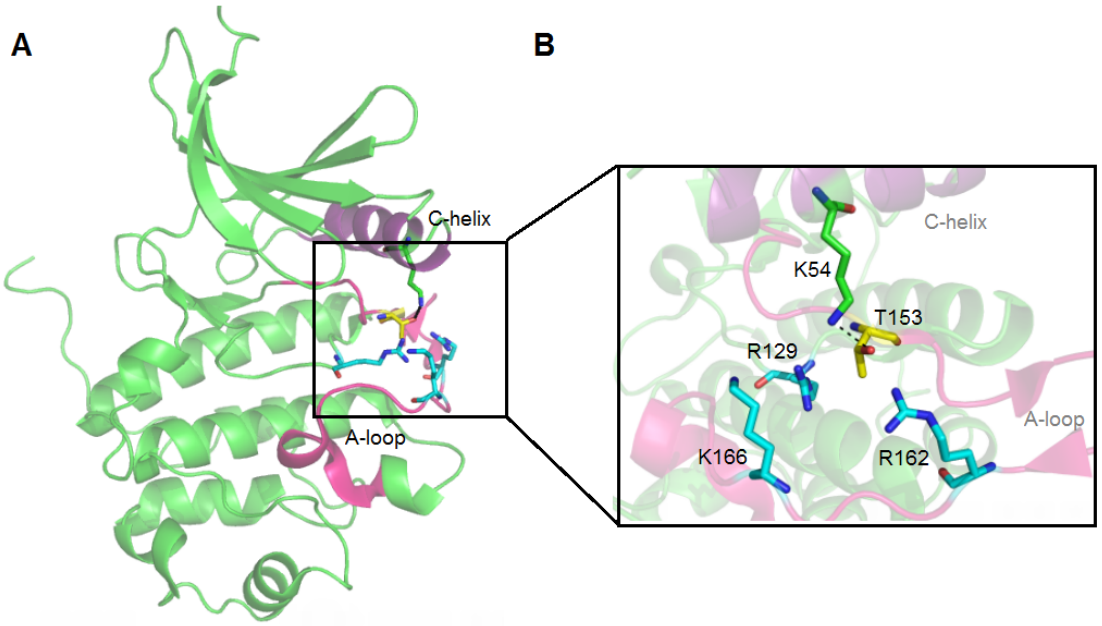
2.2.5 Non canonical-RD Pocket in Chk1

An alternative-RD pocket is identified around a putative non-consensus phosphorylation site Thr153, in A-loop of Chk1. The RMSD of the pocket from the canonical RD-pocket is 1.43 Å. The pocket is composed of the RD-Arg129 and R162, K166 (Figure 2.16 A, B) from the A-loop that are all conserved in vertebrates (Figure 2.16 C). In some lower organisms such as *S. cervisiae*, K166 is not conserved and is replaced by D166. Putative phospho-site Thr153 lies at the DFG+3 position. Presence of an alternative RD-pocket composed of invariant basic residues despite the absence of consensus-primary phosphorylation site was an intriguing feature. Our analysis suggests that Thr153 is a negative regulatory site and the alternative RD pocket locks the phosphorylated form in an inactive conformation.

So far there is no evidence of phosphorylation in the catalytic domain of Chk1(Zhao & Piwnica-Worms, 2001). Indeed, the consensus primary phosphorylation site is substituted by the basic residue K166, which is critical for substrate specificity (P. Chen *et al.*, 2000). This indicates that Chk1 is regulated by a different mechanism. In the unphosphorylated state the enzyme obtains an active conformation where T153 from the DFG+3 position forms hydrogen bonds with Lys54. Lys54 is highly conserved across the Chk1 family and is located right next to the invariant glutamate in the C-helix. The interaction between T153 & K54 is similar to the interaction between PSTAIRE

Arg50 and TPO160 (Figure 2.8 A) in CDK that “pulls” the C-helix in the active conformation.

Phosphorylation of T153 could play a negative regulatory role. Phosphorylation of T153 is likely to cause steric repulsions between T153 & K54 due to which the interaction between T153-K54 is likely to be lost. As a result, the C-helix could become disoriented, causing breakage of the crucial Lys-Glu salt-bridge. So the phosphorylation of T153 could render the kinase inactive. Upon T153 phosphorylation, the identified alternative RD-pocket is likely to adjust its position in order to act as a phospho-binding pocket and stabilize the inactive configuration. This type of regulation mechanism is reminiscent of JAK3. JAK3 has two phosphorylation sites in the activation loop, Tyr980 and Tyr981. Experimental studies have shown that phosphorylation of Tyr980 plays a positive regulatory role while that of Tyr981 plays a negative regulatory role (Zhou *et al.*, 1997). However, structural basis of these findings remain unclear. Phospho-site prediction algorithms –NETPHOS 2.0 (N Blom *et al.*, 1999), DISPHOS (Iakoucheva *et al.*, 2004) and PPSP (Xue *et al.*, 2006) do not predict Thr153 as a potential phosphosite. However our analysis suggests that T153 is a potential phosphorylation site and the impact of phosphorylation of T153 should be worth exploring.



C

```
                                                                                                K54
CHK1_Hs          WDLVQTLGEGAYGEVQ---LAVNRVTEEAVAVKIVDMKRAVDCPEN-IKKEICIN-KMLN
CHK1_Mm          WDLVQTLGEGAYGEVQ---LAVNRITEEAVAVKIVDMKRAIDCPEN-IKKEICIN-KMLS
Chk1_Sp          VELCNVLSS--CQLVK---LAVNRSTQEAVAVKILHLDKQQGAADS-VRKEVCIH-RMLG
grp_Dm           WTLAQTLGEGAYGEVK---LLINRQTGEAVAMKMVDLKKHPDAANS-VRKEVCIQ-KMLQ
CHK1_Sc          VVLGDTVGQGAFACVKNAHLQMDPSIILAVKFIHVPTCKKMGLSDKDITKEVVLQSKCSK
                   * :.:..     *:   * ::     ** .  :   :  . .:. : **: :: :

                            R129                            T153        R162  K166
CHK1_Hs          HGIG-ITHRDIKPENLLLDERDNLKISDFGLATVFRYN-NRERLLNKMCGTLPYVAPELL
CHK1_Mm          HGIG-ITHRDIKPENLLLDERDNLKISDFGLATVFRHN-NRERLLNKMCGTLPYVAPELL
Chk1_Sp          HKKG-VTHRDLKPENLLLDEHDNLKVSDFGLATVFRHQ-GKERLLGKCCGTPPYVAPEVL
grp_Dm           HQRG-IAHRDLKPENLLLDEHDNVKISDFGMATMFRCK-GKERLLDKRCGTLPYVAPEVL
CHK1_Sc          HVECGVAHRDIKPENILLDKNGNLKLADFGLASQFRRKDGTLRVSMDQRGSPPYMAPEVL
                 *    ::**:****:***:..*:*::***:*: ** : .  *:  .  *: **:***:*
```

Figure 2.16: Alternative RD-Pocket in Chk1. (A) Overall structure of CHK1 (PDB ID 2BRB) is shown in ribbon diagram. The C-helix is colored deep purple and A-loop is colored hot pink (B) Detailed view of the non-RD pocket (stick representation, colored cyan), the potential phospho-site Thr153 (stick representation, colored cyan) and Lys54 (stick representation, colored green) from the C-helix. (C) ClustalW sequence alignment of Chk1 kinase domains containing the non-RD-pocket region of *human* (Hs), *mouse* (Mm), *Sea Urchin* (Sp), *Drosophila* (Dm), *C. elegans* (Ce), *S. cerevisiae* (Sc). Basic residues forming the alternative-RD pocket are highlighted in blue. Putative phosphorylation site is highlighted in yellow. C-helix basic residue-K54 is highlighted in green.

2.3 Non-Kinase Basic Pockets

So far we have identified several functionally important non-RD pockets in different regions of protein kinases. We next wanted to investigate whether this pocket is found in other phosphorylated non-kinase proteins as well. We built a library of 469 phosphorylated non-kinase protein structures with 95% sequence identity. With canonical RD-pocket as a query the phosphorylated non-kinase protein library is searched using SPASM (Kleywegt, 1999). However, SPASM does not identify a basic pocket with the exact geometry as that of the RD-pocket around the phosphorylated sites in the non-kinase phosphorylated protein library. This suggests that the specific geometry of the RD-pocket is unique to kinases.

However, favorable charge-charge interactions have been determined at phosphorylated sites in many other non-kinase proteins, such as in phospho-peptide binding systems or at interfaces mediated by phosphorylation sites (Binding-domains, Roque, Lowe, & Cecı, 2005; Joughin, Tidor, & Yaffe, 2005; Kitchen *et al.*, 2008). We used energy analysis to identify such favorable charge-charge interactions between phospho-residue and basic residues in non-kinase phosphorylated proteins (see Methods). Phospho-groups that have high interaction with two or more basic residues (Arginine/Lysine/Histidine) were classified as phospho-sites with a basic pocket; otherwise they were classified as phospho-sites with no basic pocket. Details about how the basic pockets were identified using energy analysis are discussed in the Methods section.

Out of 469 phosphorylated non-kinase protein structures, 258 proteins had a phospho-site with a pocket, and 211 proteins had phospho-sites without a pocket. To

understand the importance of a “basic pocket” in phosphorylated proteins in general we studied the biological relevance of such pockets in a few families.

2.3.1 Basic Pocket in Glycogen Phosphorylase

A basic pocket composed of two Arginines (Arg69 & Arg43’) is identified at the homodimer interface around the key phosphorylation site Ser-14 in Glycogen phosphorylase (Figure 2.17). Glycogen phosphorylase is a prototype regulatory protein, whose activity is controlled by reversible phosphorylation and allosteric effectors (L N Johnson, 1992). It is a functional dimer and upon phosphorylation SEP-14 makes hydrogen bonds and ionic contacts to Arg69 from its own subunit, and Arg43’ from the other subunit of the dimer (L N Johnson & Barford, 1994). Phosphorylation brings about changes in the subunit-subunit interface. These changes in turn promote changes in the tertiary and quaternary conformation, coupling the event of phosphorylation with other allosteric effector sites (Sprang *et al.*, 1988). This leads to activation at the catalytic site 45Å away from the Ser-14 phosphosite (L N Johnson, 1992).

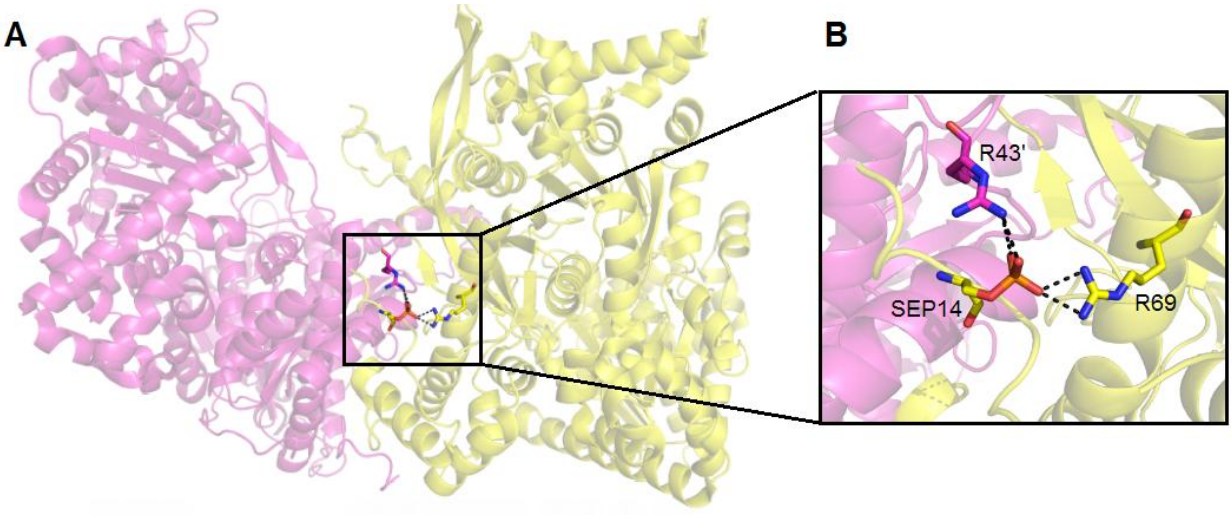


Figure 2.17: Basic pocket in glycogen phosphorylase. (A) Overall structure of glycogen phosphorylase functional dimer in R-state is shown in ribbon representation (PDB ID 1GPA). (B) Close up view of the basic pocket is shown in stick

representation. Subunit one is colored yellow and subunit two is colored magenta.

Glycogen Phosphorylase can exist in unphosphorylated b form (GPb) and phosphorylated a form (GPa) with two SEP groups while maintaining the equilibrium between two states – T & R state (Krebs & Fischer, 1956). The inactive T-state exhibits low level of affinity for substrates and an R state exhibits high affinity for substrates (Sprang *et al.*, 1988). The key difference between GPa and GPb form lies in their preference of states, GPa has a predilection for active R-state, whereas GPb has a predilection for T-state. Crystal structures of R & T state GPa and GPb have contributed a great deal to our understanding of control properties of the enzyme (Barford, Hu, & Johnson, 1991; L N Johnson, 1992; L N Johnson & Barford, 1994; Sprang *et al.*, 1988). In GPb the the 20 residue N-terminal tail obtains an irregular extended conformation, establishing intra-subunit contacts and exhibiting mobility. Upon the phosphorylation of Ser-14, N-terminal tail undergoes major conformational changes. Ser moves more than 35 Å between GPb and GPa. The N-terminal tail adopts an ordered conformation with partly 310 helix such that phosphate group establishes inter-subunit contacts. The basic pocket composed of Arg69 and Arg43' plays a crucial role in the activation mechanism by providing an appropriate environment for the dianionic phosphate. The active state conformation of the Arginine pocket would be less favorable for the unphosphorylated N-terminal tail; the position of two Arginines is indeed considerably altered upon dephosphorylation (Figure 2.18). So, the basic pocket plays a crucial role by stabilizing SEP14 and coupling the event of phosphorylation with activating conformational changes in the enzyme.

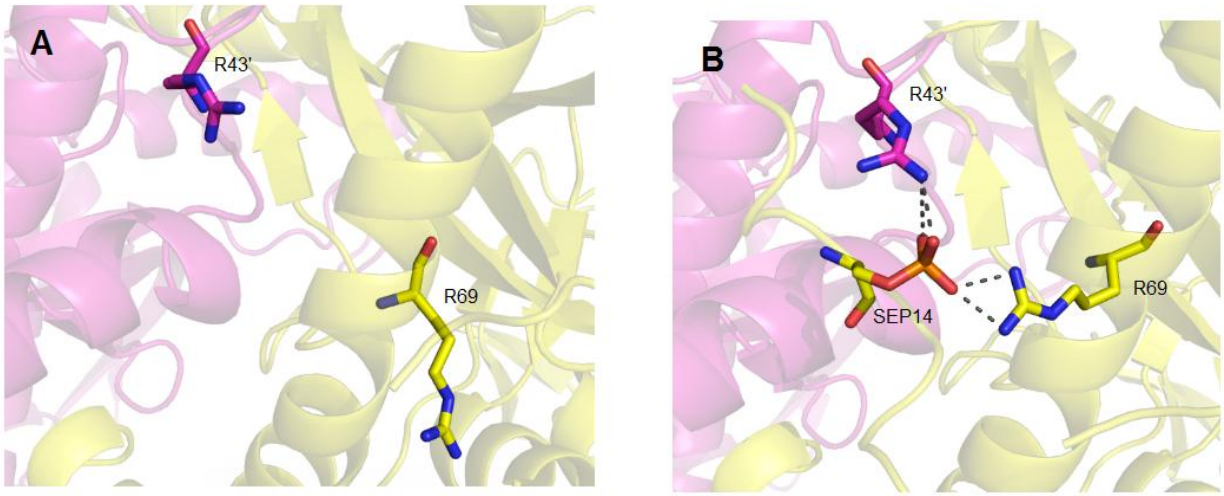


Figure 2.18: Comparison of basic pocket in phosphorylated and non-phosphorylated Glycogen Phosphorylase. (A) Structure representation of basic pocket when Ser-14 is not phosphorylated (PDB ID: 8GPB). (B) Structure representation of basic pocket when Ser-14 is phosphorylated (SEP14), PDB ID: 1GPA.

2.3.2 Basic Pocket in DEUBIQUITINASE

Another basic pocket composed of two arginines and a lysine (Arg272, Lys273, Arg74U) is found around a phosphorylated site (SEP177) in DUBA that is bound to ubiquitinase (Figure 2.19). Recently Huang et al (2012) discovered that phosphorylation is a key mode of regulation of DUBA, a deubiquitinase which belongs to the Ovarian Tumor (OTU) subfamily(O. W. Huang *et al.*, 2012). The phosphate group brackets two distant regions of the OTU domain that imprisons the C-terminal tail of ubiquitin and the basic pocket has a major contribution in this bracketing mechanism.

Ubiquitination has emerged as a common post translational modification which is crucial for protein degradation and other basic cellular processes like cell signaling & DNA damage response (Z. J. Chen & Sun, 2009; Hershko & Ciechanover, 1998). Like phosphorylation, ubiquitination is a reversible process and a versatile means of regulating protein functions. Spatial and temporal control of ubiquitination is important

for cell signaling and protein stability and this control relies on tight regulation of ubiquitin addition and removal. Deubiquitinase (Dubs), as their name suggests are ubiquitin specific proteases responsible for removing ubiquitin from target proteins. The phospho-site Ser177 responsible for activity lies in a disordered region of the protein.

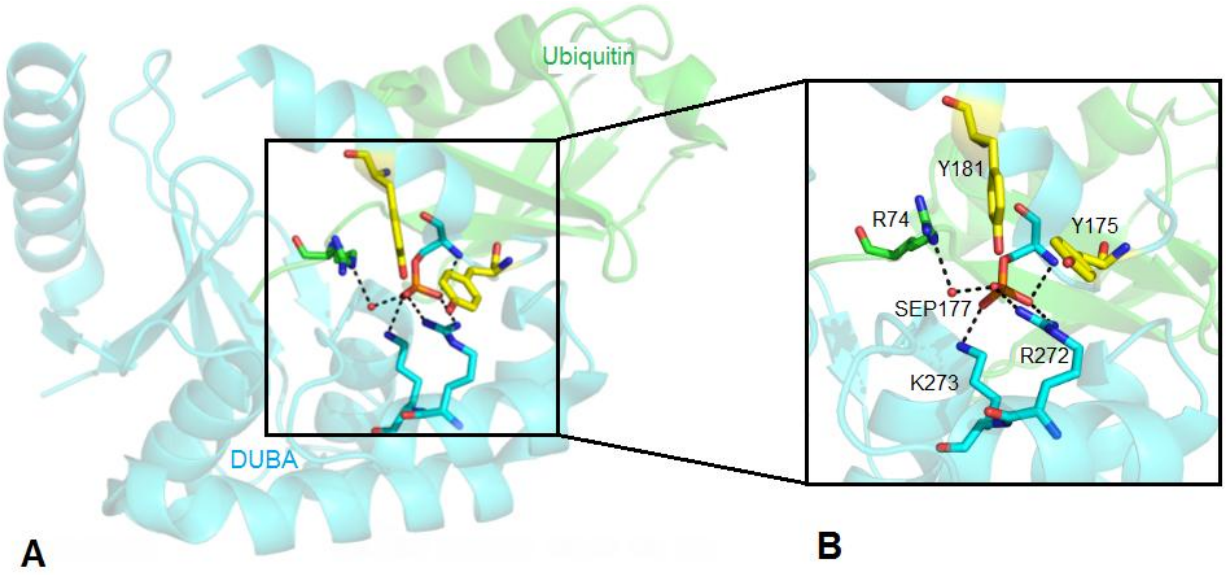


Figure 2.19: Basic pocket in Deubiquitinase. (A) Overall structure of deubiquitinase complexed with ubiquitinase shown in ribbon representation PDB ID: 3TMP. (B) Close up view of the basic pocket and phosphoresidue SEP177 shown in stick representation. Deubiquitinase is colored cyan and ubiquitinase is colored green.

While phosphorylation alone does not give rise to evident structural changes, it is crucial for ubiquitin recognition (O. W. Huang *et al.*, 2012). Binding of phosphorylated ubiquitin with DUBA induces large conformational changes in the structure (O. W. Huang *et al.*, 2012). The phosphate moiety establishes several contacts with the basic C-terminal tail of ubiquitin. The disordered region around Ser177 organizes into alpha-helix-1. Via the phosphate moiety, the alpha-helix-1 interacts with the otherwise distant central region of the OTU domain, namely alpha-helix-6. Arg272, Lys273 from alpha-helix-6 along with Arg74U (from ubiquitin) form a basic cluster around the phosphate.

Arg272 and Arg74U also form cation-pi ponds with Tyr175 and Tyr181 from alphs-helix1. This complex network of interactions between alpha-helix-1 and alpha-helix-6 clamps the C-teminal tail of ubiquitin. The induced conformational changes allow the catalytic triad to catalytically competent orientation.

2.3.3 Basic Pocket in SMAD Proteins

A basic pocket is identified at the inter-subunit interface of a phosphorylated heterotrimer complex of SMAD3-SMAD4 (Figure 2.20 A). It is observed that the phosphoresidue (SEP425) from the C-terminal tail of one subunit (SMAD-3) binds to a basic pocket (Arg378', Lys507') (Figure 2.20 B). The basic pocket is contained in the C-terminal region of the second sub-unit (SMAD4). It has been shown that phosphorylation of R-SMAD triggers the formation of the heteromeric assembly (Abdollah *et al.*, 1997), a central event in TGF-Beta signaling pathway. Phosphorylation acts as a signaling switch; phosphorylated C-terminal tail of one subunit forms specific contacts with the phosphoserine binding pocket of the neighboring subunit that subsequently leads to oligomerization (Qin *et al.*, 2001; Wu *et al.*, 2001).

SMAD proteins are intracellular mediators that transduce extracellular signals from TGF-B to the nucleus where they modulate the activity of transcription factors, their activators and repressors by direct protein-protein interaction (Attisano & Wrana, 2000; Derynck, Zhang, & Feng, 1998; Heldin, Miyazono, & Ten Dijke, 1997). The Smad protein family can be divided into three functionally distinct classes: (i) Receptor regulated Smads (R-Smads ~ Smad 1, 2, 3, 5 & 8) which are directly phosphorylated by the receptors and mediate TGF-B signaling in a pathway-specific manner, (ii) the co-mediatr Smads (Smad4) that forms heteromeric complexes with phosphorylated R-

Smads, (iii) Inhibitory Smads (Smad 6, 7) that antagonize TGF-B signaling and counteract the effect of the R-Smads (Massagué, Chen, & Massague, 2000). R-Smads and Smad4 have similar domain architecture (Heldin *et al.*, 1997)(Massague, 1998).

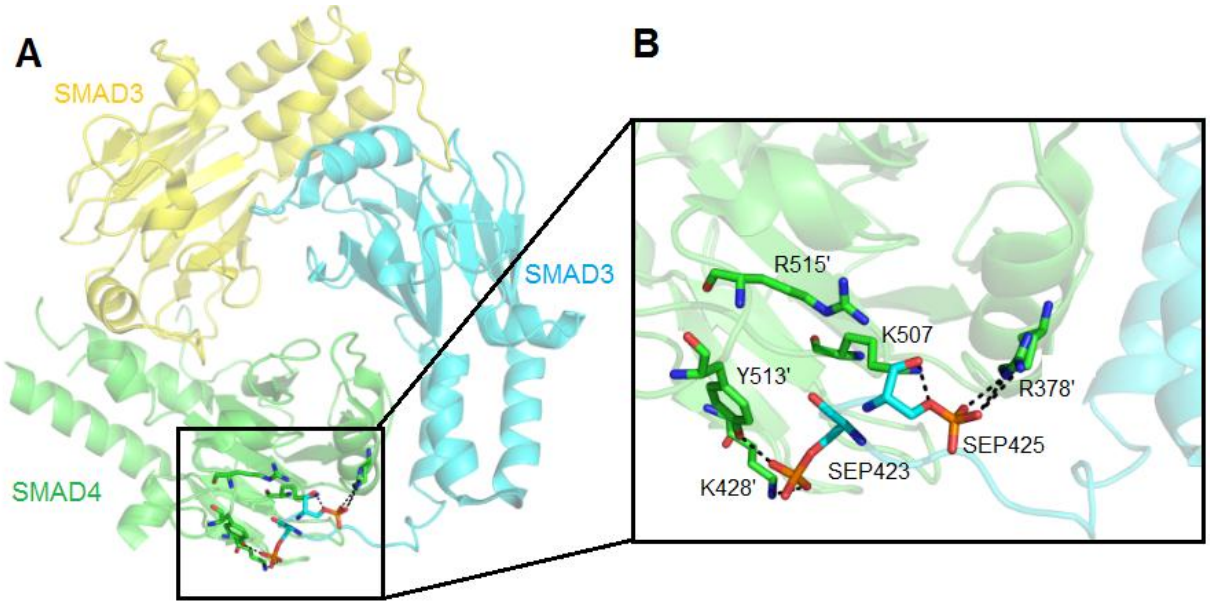


Figure 2.20: Basic pocket in SMAD proteins. (A) Overall structure of SMAD3-SMAD4 hetero-complex shown in ribbon diagram. (PDB ID 1U7F) (B) Close up view of the basic pocket shown in stick representation. Subunit 1, SMAD4 is colored green. Subunit 2, SMAD3 is colored cyan. Subunit 3, SMAD3 is colored yellow.

They have a C-terminal MH-2 domain which is responsible for oligomerization and transcriptional regulation and the N-terminal MH1 domain which is involved in DNA binding. A unique distinguishing feature of R-Smads is a conserved Ser-Ser-X-Ser motif in the C-terminal (Abdollah *et al.*, 1997; Souchelnytskyi *et al.*, 1997). Phosphorylation at the SSXS motif is indispensable in TGF-B signaling as it induces dissociation from the receptor complex as well as from accessory proteins (eg. SARA-Smad anchor for receptor activation) (Abdollah *et al.*, 1997; Souchelnytskyi *et al.*, 1997). It also allows for concomitant formation of a heterocomplex between R-Smads and Co-Smad (Smad4) (Jayaraman & Massague, 2000).

Structural studies suggest that phosphorylation of C-terminal tail functions as an assembly switch for the subunits by establishing specific contacts with the phospho-serine binding pocket of the neighboring subunit (Qin *et al.*, 2001; Wu *et al.*, 2001). The MH2 domain indeed undergoes a conformational change upon oligomerization which strengthens the hypothesis that phosphorylation of the C-terminal tail acts as a conformational switch. So, upon phosphorylation the C-terminal tail obtains an extended conformation forming specific contacts with the L3-loop B8-strand pocket of the adjacent monomer. The phosphoserines play the role of anchoring residues. The inter-subunit contacts are formed by Arg378, K507, R515, K428 along with Tyr426, which form the phosphate binding pocket. The importance of this basic pocket is reflected by the fact the phosphate binding residues are invariant among all R-Smads and Smad4 (Qin *et al.*, 2001; Wu *et al.*, 2001). Also, mutation of these residues significantly affects oligomerization (Qin *et al.*, 2001; Wu *et al.*, 2001).

2.3.4 Basic Pocket in CrkII

A basic pocket composed of two arginines (Arg20 & Arg38) was identified around phospho-residue PTR221 in CrkII proteins (Figure 2.21). Tyr221 is known to be a negative regulatory phosphorylation site (Feller, 2001; Kobashigawa *et al.*, 2007). Phosphorylation induces large conformational changes that lock the protein in a ligand-binding incompetent conformation. The basic pocket behaves as a phosphate recognition site in the SH2 domain and the intramolecular interactions between the basic pocket and PTR22 contributes to the autoinhibitory mechanism (Feller, 2001; Kobashigawa *et al.*, 2007).

CRK-II proteins are archetypal signaling adaptor proteins that link Tyrosine Kinases to small G proteins (Feller, 2001). The CRK-II proteins comprise of an N-terminal Src Homology 1 domain and two Src Homology 3 domains. To function as adaptor proteins, SH2 and n-SH3 domains act as molecular adhesives that capture diverse cellular signals and accordingly regulate the signal transduction pathways spatially and temporally. The SH2 domain mediates association with tyrosine phosphorylated receptors and the N-terminal SH3 (n-SH3) domain is responsible for transmitting signals to small G-proteins (Feller, Knudsen, & Hanafusa, 1994). The C-terminal SH3 (c-SH3) domain negatively regulates the biological activity of n-SH3 (Feller, 2001)(Kobashigawa *et al.*, 2007). The two SH3 domains are tethered together by a 50 residue long linker that contains a phophorylatable tyrosine (Tyr221).

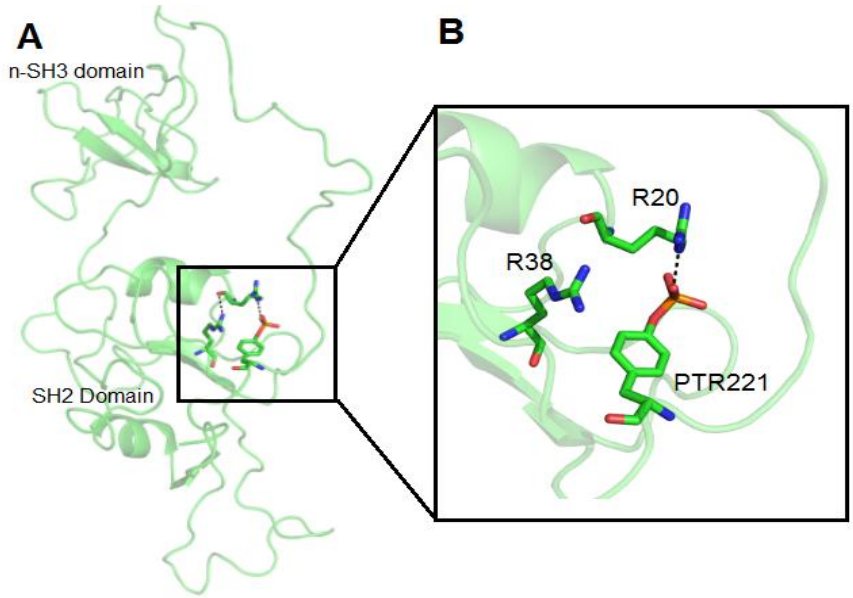


Figure 2.21: Basic Pocket in CrkII. (A) Overall structure of CrkII shown in ribbon representation (PDB ID 2DVJ). (B) Close up view of the basic pocket and the phospho-residue PTR221 shown in stick representation.

Elucidation of the three dimensional solution structure of CRK-II in the unphosphorylated and phosphorylated forms determined by NMR spectroscopy verified that phosphorylation exerted allosteric conformational changes (Kobashigawa *et al.*, 2007). Upon phosphorylation of Y221, the Crk-II adapter proteins undergo a large conformational change, binding the phosphosite to its own SH2 domain. This intramolecular reorganization induced by phosphorylation gives rise to an autoinhibitory conformation that prevents binding of physiological ligands. The phospho-tyrosine interacts with two Arginines (Arg38 and Arg20) from the SH2 domain that forms a basic phospho-recognizing pocket. The electrostatic interactions between the phosphotyrosine and the basic pocket facilitate intramolecular domain-domain interactions (Kobashigawa *et al.*, 2007). In the unphosphorylated form, the SH2 domain does not interact with the linker region and the Crk-II obtains an uninhibited, ligand-binding-competent conformation (Kobashigawa *et al.*, 2007).

In the above examples basic pocket is undoubtedly a phosphate recognition site that is very highly conserved across the given family. The basic pocket establishes numerous well defined interactions with the phospho-residue via hydrogen bonds, or ionic side-chain interactions. The analysis suggests that the basic pocket is a feature of phosphorylation sites that induce conformation changes. The main function of the pocket is to allosterically couple the phosphorylation site to distal active and regulatory sites. Unlike kinases, the geometry of the basic pocket is not conserved around regulatory phospho-sites in other proteins. Nonetheless, a pocket in general is conserved. Identification of such pockets combined with simulation studies can assist

analysis aimed to associate the event of phosphorylation with conformational change as the target phosphate binding area is already identified.

CHAPTER 3

CONCLUSIONS

We have studied various canonical and alternate RD-pockets around the consensus phosphorylation site in the A-loop of protein kinases. Our results show that although the RD-pocket is not conserved sequentially, it is conserved structurally. The RD-pocket is a spatial motif that is conserved across families that require activation loop phosphorylation. Different families have evolved diverse RD-pockets in order to couple their unique regulatory mechanisms with the event of activation loop phosphorylation. We also studied the prevalence of the RD-pocket in other regions of the kinase that do not encompass the activation-loop consensus phosphorylation site. Our results indicate that several non-RD pockets exist around regulatory phosphorylation site other than that in the A-loop. We further explored the possible biological relevance of a few non-RD pockets which reveals that the basic pocket is phospho- recognition site associated with charge-charge interaction and conformational changes. Our studies suggest that identification of such pockets can be used to identify potential allosteric sites that are regulated by phosphorylation.

CHAPTER 4

METHODS

4.1 Sequence Data

Sequences used consist of 13,346 protein kinase sequences retrieved from NCBI-nr of July 2010. Subgroup specific distribution of the 13346 sequences is as follows - 1449 AGC, 2942 CAMK, 437 CK1, 1721 CMGC, 1147 STE, 1166 TK, 1902 TKL and 2582 Other Kinase sequences.

4.2 Multiple Sequence Alignment

A curated set of ePK family profiles are constructed using previously annotated sequences from diverse model organisms. The strategy for classifying these profiles is based on previous kinase studies. (Hanks & Hunter, 1995), (Manning *et al.*, 2002). MAPGAPS (Neuwald, 2009) program is used in conjunction with the curated profile sets to identify, classify and align the protein kinase sequences. MAPGAPS (Neuwald, 2009) first identifies the kinase domain using the key motifs and then it assigns each sequence to the best-matching family in the query profile. These sequences are then aligned to the kinase consensus sequence and the fragmentary sequences are deleted. The multiple alignments illustrated in figures are obtained using CLUSTALW (Larkin *et al.*, 2007).

4.3 Residue Conservation

Residue conservation is counted from the final alignment obtained from MAPGAPS (Neuwald, 2009) using a custom script.

4.4 Structural Conservation

SPASM (Kleywegt, 1999), a program that recognizes spatial motifs is used to identify the RD pockets in the different kinase families. A kinase library is built that contains 1924 kinase structures retrieved from the Protein Data Bank. A representative library is built by carefully selecting 160 non-redundant structures from 137 kinase families that are preferably in the active state conformation.

In both the kinase libraries, coordinates of all phosphoresidues i.e. TPO, SEP and PTR are replaced by Thr, Ser and Tyr respectively because SPASM does not identify “HETATM” when looking for motifs. The canonical RD pocket in CDK (PDB ID: 1QMZ) composed of Arg 50, Arg126, Arg50 & Thr160 (TPO160 →Thr160) is used as a query motif. The maximum superposition RMSD parameter is adjusted from 1A -2A to identify canonical, alternate and non-RD pockets in the kinase structures. User defined substitutions allow the identification of pockets that can have a Lys or Histidine in place of Arginine and Ser or Tyr in place of Thr in the canonical RD-pocket.

4.5 Analysis of Energy of Interactions

The energy of a residue-residue interaction is calculated as the enthalpy of interaction and not necessarily the free energy of interaction. The partial charges and van der Waals radii for residues are assigned using the program PDB2PQR using the AMBER forcefield. The final equation used for the enthalpy is as follows:

$$\text{Energy} = \sum_{i<j}^{atoms}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + \sum_{i<j}^{atoms}\frac{q_i q_j}{\varepsilon r_{ij}}.$$

The van der Waals parameters Aij and Bij for each atom type pairs were taken from AMBER forcefield. Surface accessibility is calculated for every residue which is used to calculate the electrostatic energy values. The dielectric constant used in calculation of electrostatic energy is linearly scaled between 2.0 for protein interiors (sum of surface accessibility < 5%), and 80.0 for surface accessible atom pairs. Such a method has been implemented in PropKa, and has been shown to improve estimation of electrostatic energies. The total energy of interaction between two residues is taken as the summation over energies of all atoms within the two residues. Consecutive residues are eliminated from analysis.

4.6 Identification of Basic Pockets in Proteins

PDB search tool is used to generate a list of proteins with either SEP, TPO or PTR listed as a part of the PDB. A redundancy cutoff of 95% is applied to obtain non-redundant homologs for each protein. The calculation of enthalpy is carried out as given in the previous sub-section. Modified force-field parameters for SEP, TPO and PTR from Bryce group (http://www.pharmacy.manchester.ac.uk/bryce/amber/) are used in PDB2PQR to calculate enthalpy of interaction for these residues.

For every phosphorylated residue, basic residues (Arg, Lys or His) within 10 Angstrom radius are chosen as interacting partner if the energy of interaction is less than -5 kcal/mol. From the dataset obtained, phospho-residues that showed no basic residues or only one basic residue around the phosphorylated residue are classified as a phospho-group without a pocket. If two or more basic residues are seen in a high energy interaction, the phospho-group is classified as situated within a pocket.

REFERENCES

Abdollah, S., Macías-Silva, M., Tsukazaki, T., Hayashi, H., Attisano, L., & Wrana, J. L. (1997). TbetaRI phosphorylation of Smad2 on Ser465 and Ser467 is required for Smad2-Smad4 complex formation and signaling. *The Journal of Biological Chemistry*, *272*(44), 27678-27685. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9346908

Abe, M. K., Kuo, W.-liang, Hershenson, M. B., & Rosner, M. R. (1999). (ERK7), a Novel ERK with a C-Terminal Domain That Regulates Its Activity , Its Cellular Localization , and Cell Growth Extracellular Signal-Regulated Kinase 7 (ERK7), a Novel ERK with a C-Terminal Domain That Regulates Its Activity , Its Cellular Loca, *7*.

Aberg, E., Torgersen, K. M., Johansen, B., Keyse, S. M., Perander, M., & Seternes, O.-M. (2009). Docking of PRAK/MK5 to the Atypical MAPKs ERK3 and ERK4 Defines a Novel MAPK Interaction Motif. *The Journal of Biological Chemistry*, *284*(29), 19392-19401. American Society for Biochemistry and Molecular Biology. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2740564&tool=pmcentrez&rendertype=abstract

Attisano, L., & Wrana, J. L. (2000). Smads as transcriptional co-modulators. *Current Opinion in Cell Biology*, *12*(2), 235-243. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10712925

Barford, D., Hu, S. H., & Johnson, L. N. (1991). Structural mechanism for glycogen phosphorylase control by phosphorylation and AMP. *Journal of Molecular Biology*, *218*(1), 233-260. Elsevier. Retrieved from http://www.sciencedirect.com/science/article/pii/002228369190887C

Bayliss, R., Sardon, T., Vernos, I., & Conti, E. (2003). Structural basis of Aurora-A activation by TPX2 at the mitotic spindle. *Molecular cell*, *12*(4), 851-62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14580337

Binding-domains, P., Roque, A., Lowe, C. R., & Cecı, A. (2005). Lessons From Nature : On the Molecular Recognition Elements of the. *Biotechnology*. doi:10.1002/bit.20561

Blanco, S., Klimcakova, L., Vega, F. M., & Lazo, P. A. (2006). The subcellular localization of vaccinia-related kinase-2 (VRK2) isoforms determines their different effect on p53 stability in tumour cell lines. *The FEBS journal*, *273*(11), 2487-2504. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16704422

Blom, N, Gammeltoft, S., & Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of Molecular Biology*, *294*(5), 1351-1362. Elsevier. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10600390

Blom, Nikolaj, Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S., & Brunak, S. (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*, *4*(6), 1633-1649. Mendeley Ltd. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15174133

Buschbeck, M., & Ullrich, A. (2005). The unique C-terminal tail of the mitogen-activated protein kinase ERK5 regulates its activation and nuclear shuttling. *The Journal of biological chemistry*, *280*(4), 2659-67. doi:10.1074/jbc.M412599200

Chen, H., Ma, J., Li, W., Eliseenkova, A. V., Xu, C., Neubert, T. a, Miller, W. T., *et al.* (2007). A molecular brake in the kinase hinge region regulates the activity of receptor tyrosine kinases. *Molecular cell*, *27*(5), 717-30. doi:10.1016/j.molcel.2007.06.028

Chen, P., Luo, C., Deng, Y., Ryan, K., Register, J., Margosiak, S., Tempczyk-russell, A., *et al.* (2000). ˚ Crystal Structure of Human Cell Cycle The 1 . 7 A Checkpoint Kinase Chk1 : Implications for Chk1 Regulation 3565 General Atomics Court, *100*, 681-692.

Chen, Z. J., & Sun, L. J. (2009). Nonproteolytic functions of ubiquitin in cell signaling. *Molecular Cell*, *33*(3), 275-286. Elsevier Inc. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19217402

Cowan-Jacob, S. W., Fendrich, G., Manley, P. W., Jahnke, W., Fabbro, D., Liebetanz, J., & Meyer, T. (2005). The crystal structure of a c-Src complex in an active conformation suggests possible steps in c-Src activation. *Structure London England 1993*, *13*(6), 861-871. Elsevier Ltd. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15939018

Dar, A. C., Dever, T. E., & Sicheri, F. (2005). Higher-order substrate recognition of eIF2alpha by the RNA-dependent protein kinase PKR. *Cell*, *122*(6), 887-900. doi:10.1016/j.cell.2005.06.044

Derynck, R., Zhang, Y., & Feng, X. H. (1998). Smads: transcriptional activators of TGF-beta responses. *Cell*, *95*(6), 737-740. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9865691

Dickens, M., Rogers, J. S., Cavanagh, J., Raitano, A., Xia, Z., Halpern, J. R., Greenberg, M. E., *et al.* (1997). A cytoplasmic inhibitor of the JNK signal transduction pathway. *Science*, *277*(5326), 693-696. American Association for the Advancement of Science. Retrieved from http://www.sciencemag.org/cgi/doi/10.1126/science.277.5326.693

Diella, F., Cameron, S., Gemünd, C., Linding, R., Via, A., Kuster, B., Sicheritz-Pontén, T., *et al.* (2004). Phospho.ELM: A database of experimentally verified phosphorylation sites in eukaryotic proteins. *BMC Bioinformatics*, *5*(1), 79. BioMed Central. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15212693

Durek, P., Schudoma, C., Weckwerth, W., Selbig, J., & Walther, D. (2009). **Detection and characterization of 3D-signature phosphorylation site motifs and their contribution towards improved phosphorylation site prediction in proteins. *BMC bioinformatics*, *10*, 117. doi:10.1186/1471-2105-10-117

Eyers, P. A., Erikson, E., Chen, L. G., & Maller, J. L. (2003). A Novel Mechanism for Activation of the Protein Kinase Aurora A. *Current*, *13*(8), 691-697. Elsevier. doi:10.1016/S

Fan, S.-cai, & Zhang, X.-gong. (2005). **Characterizing the Microenvironment Surrounding Phosphorylated Protein Sites, *3*(4), 213-217.

Feller, S. M. (2001). Crk family adaptors-signalling complex formation and biological roles. *Oncogene*, *20*(44), 6348-6371. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11607838

Feller, S. M., Knudsen, B., & Hanafusa, H. (1994). c-Abl kinase regulates the protein binding activity of c-Crk. *the The European Molecular Biology Organization Journal*, *13*(10), 2341-2351. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=395099&tool=pmcentrez&rendertype=abstract

Fischer, E. H., & Krebs, E. G. (1955). Conversion of phosphorylase b to phosphorylase a in muscle extracts. *The Journal of Biological Chemistry*, *216*(1), 121-132. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/13252012

Fitch, I., Dahmann, C., Surana, U., Amon, A., Nasmyth, K., Goetsch, L., Byers, B., *et al.* (1992). Characterization of four B-type cyclin genes of the budding yeast Saccharomyces cerevisiae. *Molecular Biology of the Cell*, *3*(7), 805-818. Retrieved from

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=275636&tool=pmcentrez&rendertype=abstract

Gnad, F., Ren, S., Cox, J., Olsen, J. V., Macek, B., Oroshi, M., & Mann, M. (2007). PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biology*, *8*(11), R250. BioMed Central. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2258193&tool=pmcentrez&rendertype=abstract

Greenman, C., Stephens, P., Smith, R., Dalgliesh, G. L., Hunter, C., Bignell, G., Davies, H., *et al.* (2007). Patterns of somatic mutation in human cancer genomes. *Nature*, *446*(7132), 153-158. Nature Publishing Group. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17344846

Haase, S. B., Winey, M., & Reed, S. I. (2001). Multi-step control of spindle pole body duplication by cyclin-dependent kinase. *Nature Cell Biology*, *3*(1), 38-42. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11146624

Hanks, S. K., & Hunter, T. (1995). The eukaryotic protein kinase superfamily : (catalytic) domam structure and classification of the. *FASEB*.

Heldin, C. H., Miyazono, K., & Ten Dijke, P. (1997). TGF-beta signalling from cell membrane to nucleus through SMAD proteins. *Nature*, *390*(6659), 465-471. Macmillan Magazines Ltd. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9393997

Heo, Y.-S., Kim, S.-K., Seo, C. I., Kim, Y. K., Sung, B.-J., Lee, H. S., Lee, J. I., *et al.* (2004). Structural basis for the selective inhibition of JNK1 by the scaffolding protein JIP1 and SP600125. *The EMBO journal*, *23*(11), 2185-95. doi:10.1038/sj.emboj.7600212

Hershko, A., & Ciechanover, A. (1998). The ubiquitin system. *Annual Review of Biochemistry*, *67*(10), 425-79. Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA. doi:10.1146/annurev.biochem.67.1.425

Huang, H.-D., Lee, T.-Y., Tzeng, S.-W., & Horng, J.-T. (2005). KinasePhos: a web tool for identifying protein kinase-specific phosphorylation sites. *Nucleic Acids Research*, *33*(Web Server issue), W226-W229. Oxford University Press. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15980458

Huang, O. W., Ma, X., Yin, J., Flinders, J., Maurer, T., Kayagaki, N., Phung, Q., *et al.* (2012). Phosphorylation-dependent activity of the deubiquitinase DUBA. *Nature structural & molecular biology*, *19*(2), 171-5. doi:10.1038/nsmb.2206

Hubbard, S. R. (1997). Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. *The EMBO journal*, *16*(18), 5572-81. doi:10.1093/emboj/16.18.5572

Hubbard, S. R., Wei, L., Ellis, L., & Hendrickson, W. A. (1994). Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature*, *372*(6508), 746-754. doi:10.1038/372746a0

Huse, M., & Kuriyan, J. (2002). The conformational plasticity of protein kinases. *Cell*, *109*(3), 275-82. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12015977

Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z., & Dunker, A. K. (2004). The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Research*, *32*(3), 1037-1049. Oxford University Press. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14960716

Jayaraman, L., & Massague, J. (2000). Distinct Oligomeric States of SMAD Proteins in the Transforming Growth Factor- Pathway. *Biochemistry*, *275*(52), 40710-40717. doi:10.1074/jbc.M005799200

Johnson, L N. (1992). Glycogen phosphorylase: control by phosphorylation and allosteric effectors. *The FASEB journal official publication of the Federation of American Societies for Experimental Biology*, *6*(6), 2274-2282. FASEB. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1544539

Johnson, L N, & Barford, D. (1994). Electrostatic effects in the control of glycogen phosphorylase by phosphorylation. *Protein Science*, *3*(10), 1726-1730. Cold Spring Harbor Laboratory Press.

Johnson, L N, Noble, M. E., & Owen, D. J. (1996). Active and inactive protein kinases: structural basis for regulation. *Cell*, *85*(2), 149-158. Elsevier Inc. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8612268

Johnson, Louise N. (2009). Protein kinase inhibitors: contributions from structure to clinical compounds. *Quarterly reviews of biophysics*, *42*(1), 1-40. doi:10.1017/S0033583508004745

Johnson, Louise N, & Lewis, R. J. (2001). Structural Basis for Control by Phosphorylation.

Joughin, B. A., Tidor, B., & Yaffe, M. B. (2005). A computational method for the analysis and prediction of protein : phosphopeptide-binding sites. *Protein Science*, *14*(1), 131-139. Wiley Online Library. doi:10.1110/ps.04964705.mains

Kaplan, K. B., Bibbins, K. B., Swedlow, J. R., Arnaud, M., Morgan, D. O., & Varmus, H. E. (1994). Association of the amino-terminal half of c-Src with focal adhesions

alters their properties and is regulated by phosphorylation of tyrosine 527. *the The European Molecular Biology Organization Journal*, *13*(20), 4745-4756. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=395413&tool=pmcentrez&rendertype=abstract

Keaton, M. A., Bardes, E. S. G., Marquitz, A. R., Freel, C. D., Zyla, T. R., Rudolph, J., & Lew, D. J. (2007). Differential susceptibility of yeast S and M phase CDK complexes to inhibitory tyrosine phosphorylation. *Current Biology*, *17*(14), 1181-1189. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2034293&tool=pmcentrez&rendertype=abstract

Kentrup, H., Becker, W., Heukelbach, J., Wilmes, A., Schürmann, A., Huppertz, C., Kainulainen, H., *et al.* (1996). Dyrk, a dual specificity protein kinase with unique structural features whose activity is dependent on tyrosine residues between subdomains VII and VIII. *The Journal of Biological Chemistry*, *271*(7), 3488-3495. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8631952

Kitchen, J., Saunders, R. E., & Warwicker, J. (2008). **Charge environments around phosphorylation sites in proteins. *BMC structural biology*, *8*, 19. doi:10.1186/1472-6807-8-19

Kleywegt, G. J. (1999). Recognition of spatial motifs in protein structures. *Journal of molecular biology*, *285*(4), 1887-97. doi:10.1006/jmbi.1998.2393

Knighton, D. R., Zheng, J. H., Ten Eyck, L. F., Ashford, V. A., Xuong, N. H., Taylor, S. S., & Sowadski, J. M. (1991). Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science*, *253*(5018), 407-414. American Association for the Advancement of Science. doi:10.1126/science.1862342

Knighton, D. R., Zheng, J. H., Ten Eyck, L. F., Xuong, N. H., Taylor, S. S., & Sowadski, J. M. (1991). Structure of a peptide inhibitor bound to the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science*, *253*(5018), 414-420. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=1862343

Kobashigawa, Y., Sakai, M., Naito, M., Yokochi, M., Kumeta, H., Makino, Y., Ogura, K., *et al.* (2007). Structural basis for the transforming activity of human cancer-related signaling adaptor protein CRK. *Nature structural & molecular biology*, *14*(6), 503-10. doi:10.1038/nsmb1241

Krebs, E. G. (1998). An accidental biochemist. *Annual Review of Biochemistry*, *67*(1), xii-xxxii. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9759479

Krebs, E. G., & Fischer, E. H. (1956). The phosphorylase b to a converting enzyme of rabbit skeletal muscle. *Biochimica et Biophysica Acta*, *20*(1), 297-301. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=13315361

Kufer, T. A., Silljé, H. H. W., Körner, R., Gruss, O. J., Meraldi, P., & Nigg, E. A. (2002). Human TPX2 is required for targeting Aurora-A kinase to the spindle. *The Journal of Cell Biology*, *158*(4), 617-623. The Rockefeller University Press. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2174010&tool=pmcentrez&rendertype=abstract

Lacey, K. R., Jackson, P. K., & Stearns, T. (1999). Cyclin-dependent kinase control of centrosome duplication. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(6), 2817-2822. The National Academy of Sciences. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19836237

Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan Pa McWilliam, H., Valentin, F., Wallace, I., *et al.* (2007). ClustalW and ClustalX version 2 (2007). *Bioinformatics*, *23*, 2947-2948.

Lewis, T. S., Shapiro, P. S., & Ahn, N. G. (1998). Signal transduction through MAP kinase cascades. *Advances in Cancer Research*, *74*(49-139), 49-139. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=9561267

Li, M., Liu, J., & Zhang, C. (2011). Evolutionary history of the vertebrate mitogen activated protein kinases family. *PloS one*, *6*(10), e26999. doi:10.1371/journal.pone.0026999

Lim, H. H., Goh, P. Y., & Surana, U. (1996). Spindle pole body separation in Saccharomyces cerevisiae requires dephosphorylation of the tyrosine 19 residue of Spindle Pole Body Separation in Saccharomyces cerevisiae Requires Dephosphorylation of the Tyrosine 19 Residue of Cdc28.

Littlepage, L. E., Wu, H., Andresson, T., Deanehan, J. K., Amundadottir, L. T., & Ruderman, J. V. (2002). Identification of phosphorylated residues that affect the activity of the mitotic kinase Aurora-A. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(24), 15440-15445. National Academy of Sciences. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=137735&tool=pmcentrez&rendertype=abstract

Manning, G., Whyte, D. B., Martinez, R., Hunter, T., & Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science (New York, N.Y.)*, *298*(5600), 1912-34. doi:10.1126/science.1075762

Massague, J. (1998). TGF-beta signal transduction. *Annual Review of Biochemistry*, *67*, 753-791. doi:10.1146/annurev.biochem.67.1.753

Massagué, J., Chen, Y.-guang, & Massague, J. (2000). Controlling TGF- β signaling. *Genes Development*, *14*(6), 627-644. Cold Spring Harbor Lab. doi:10.1101/gad.14.6.627

Mayans, O., Van Der Ven, P. F., Wilm, M., Mues, A., Young, P., Fürst, D. O., Wilmanns, M., *et al.* (1998). Structural basis for activation of the titin kinase domain during myo ® brillogenesis correction Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen. *Nature*, *395*(6705), 863-9. doi:10.1038/27603

Min, X., Lee, B.-H., Cobb, M. H., & Goldsmith, E. J. (2004). Crystal structure of the kinase domain of WNK1, a kinase that causes a hereditary form of hypertension. *Structure*, *12*(7), 1303-11. doi:10.1016/j.str.2004.04.014

Mohammadi, M., Schlessinger, J., & Hubbard, S. R. (1996). Structure of the FGF receptor tyrosine kinase domain reveals a novel autoinhibitory mechanism. *Cell*, *86*(4), 577-87. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8752212

Morgan, D. O. (1997). Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annual Review of Cell and Developmental Biology*, *13*(1), 261-291. Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9442875

Murray, A. W. (2004). Recycling the cell cycle: cyclins revisited. *Cell*, *116*(2), 221-234. Elsevier. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14744433

Neuwald, A. F. (2009). Rapid detection, classification and accurate alignment of up to a million or more related protein sequences. *Bioinformatics (Oxford, England)*, *25*(15), 1869-75. doi:10.1093/bioinformatics/btp342

Nichols, R. J., Wiebe, M. S., & Traktman, P. (2006). The vaccinia-related kinases phosphorylate the N' terminus of BAF, regulating its interaction with DNA and its retention in the nucleus. (A. Gregory Matera, Ed.)*Molecular Biology of the Cell*, *17*(5), 2451-2464. The American Society for Cell Biology. Retrieved from http://www.molbiolcell.org/cgi/content/abstract/17/5/2451

Nolen, B., Taylor, S., & Ghosh, G. (2004). Regulation of protein kinases; controlling activity through activation segment conformation. *Molecular cell*, *15*(5), 661-75. doi:10.1016/j.molcel.2004.08.024

Obenauer, J. C., Cantley, L. C., & Yaffe, M. B. (2003). Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Research*, *31*(13), 3635-3641. Oxford University Press. Retrieved from http://www.nar.oupjournals.org/cgi/doi/10.1093/nar/gkg584

Oppermann, F. S., Gnad, F., Olsen, J. V., Hornberger, R., Greff, Z., Kéri, G., Mann, M., *et al.* (2009). Large-scale Proteomics Analysis of the Human Kinome. *Molecular cellular proteomics MCP*, *8*(7), 1751-1764. The American Society for Biochemistry and Molecular Biology. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19369195

Pinna, L. A., Ruzzene, M., Biologica, C., & Padova, U. (1996). How do protein kinases recognize their substrates? *Biomembranes*, *1314*(3), 191-225. Elsevier. doi:10.1016/S0167-4889(96)00083-3

Plewczyński, D., Tkacz, A., Godzik, A., & Rychlewski, L. (2005). A support vector machine approach to the identification of phosphorylation sites. *Cellular molecular biology letters*, *10*(1), 73-89. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.108.6083&rep=rep1&type=pdf

Qin, B. Y., Chacko, B. M., Lam, S. S., Caestecker, M. P. D., Correia, J. J., & Lin, K. (2001). Structural Basis of Smad1 Activation by Receptor Kinase Phosphorylation, *8*, 1303-1312.

Richardson, H., Lew, D. J., Henze, M., Sugimoto, K., & Reed, S. I. (1992). Cyclin-B homologs in Saccharomyces cerevisiae function in S phase and in G2. *Genes & Development*, *6*(11), 2021-2034. Retrieved from http://www.genesdev.org/cgi/doi/10.1101/gad.6.11.2021

Rikova, K., Guo, A., Zeng, Q., Possemato, A., Yu, J., Haack, H., Nardone, J., *et al.* (2007). Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell*, *131*(6), 1190-1203. Elsevier Inc. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18083107

Robinson, M. J., & Cobb, M. H. (1997). Mitogen-activated protein kinase pathways. *Current Opinion in Cell Biology*, *9*(2), 180-186. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9069255

Roskoski, R. (2004). Src protein-tyrosine kinase structure and regulation. *Biochemical and Biophysical Research Communications*, *324*(4), 1155-1164. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15504335

Sanz-García, M., López-Sánchez, I., & Lazo, P. A. (2008). Proteomics Identification of Nuclear Ran GTPase as an Inhibitor of Human VRK1 and VRK2 (Vaccinia-related Kinase) Activities*S. *Molecular cellular proteomics MCP*, *7*(11), 2199-2214.

American Society for Biochemistry and Molecular Biology. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2577208&tool=pmcentrez&rendertype=abstract

Scheeff, E. D., Eswaran, J., Bunkoczi, G., Knapp, S., & Manning, G. (2009). Structure of the Pseudokinase VRK3 Reveals a Degraded Catalytic Site, a Highly Conserved Kinase Fold, and a Putative Regulatory Binding Site. *Structure London England 1993*, *17*(1), 128-138. Cell Press. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2639636&tool=pmcentrez&rendertype=abstract

Schmelzle, K., & White, F. M. (2006). Phosphoproteomic approaches to elucidate cellular signaling networks. *Current Opinion in Biotechnology*, *17*(4), 406-14. Elsevier. doi:10.1016/j.copbio.2006.06.004

Sicheri, F., Moarefi, I., & Kuriyan, J. (1997). Crystal structure of the Src family tyrosine kinase Hck. *Nature*, *385*(6617), 602-609. Nature Publishing Group. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9024658

Souchelnytskyi, S., Tamaki, K., Engström, U., Wernstedt, C., Ten Dijke, P., & Heldin, C. H. (1997). Phosphorylation of Ser465 and Ser467 in the C terminus of Smad2 mediates interaction with Smad4 and is required for transforming growth factor-beta signaling. *The Journal of Biological Chemistry*, *272*(44), 28107-28115. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9346966

Sprang, S. R., Acharya, K. R., Goldsmith, E. J., Stuart, D. I., Varvill, K., Fletterick, R. J., Madsen, N. B., *et al.* (1988). Structural changes in glycogen phosphorylase induced by phosphorylation. *Nature*, *336*(6196), 215-221. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3194008

Superti-Furga, G., & Courtneidge, S. A. (1995). Structure-function relationships in Src family and related protein tyrosine kinases. *BioEssays news and reviews in molecular cellular and developmental biology*, *17*(4), 321-330. Wiley Online Library. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7537961

Walter, A. O., Seghezzi, W., Korver, W., Sheung, J., & Lees, E. (2000). The mitotic serine/threonine kinase Aurora2/AIK is regulated by phosphorylation and degradation. *Oncogene*, *19*(42), 4906-4916. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11039908

Wenqing, X., & Harrison, S. (1997). Three-dimensional structure of the tyrosine kinase c-Src. *Nature*, *385*, 595-602. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Three-dimensional+structure+of+the+tyrosine+kinase+c-Src#0

Williams, J. C., Weijland, A., Gonfloni, S., Thompson, A., Courtneidge, S. A., Superti-Furga, G., & Wierenga, R. K. (1997). The 2.35 Å\ crystal structure of the inactivated form of chicken Src: A dynamic molecule with multiple regulatory interactions. *Journal of Molecular Biology*, *274*(5), 757-775. Retrieved from http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/referer?http://www.idealibrary.com/cgi-bin/links/citation/0022-2836/274/757

Wissing, J., Jänsch, L., Nimtz, M., Dieterich, G., Hornberger, R., Kéri, G., Wehland, J., *et al.* (2007). Proteomics analysis of protein kinases by target class-selective prefractionation and tandem mass spectrometry. *Molecular cellular proteomics MCP*, *6*(3), 537-47. doi:10.1074/mcp.T600062-MCP200

Wu, J.-wei, Hu, M., Chai, J., Seoane, J., Huse, M., Li, C., Rigotti, D. J., *et al.* (2001). Crystal Structure of a Phosphorylated Smad2 : Recognition of Phosphoserine by the MH2 Domain and Insights on Smad Function in TGF- ᴺʟ Signaling, *8*, 1277-1289.

Xue, Y., Li, A., Wang, L., Feng, H., & Yao, X. (2006). PPSP: prediction of PK-specific phosphorylation site with Bayesian decision theory. *BMC Bioinformatics*, *7*, 163. BioMed Central. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1435943&tool=pmcentrez&rendertype=abstract

Yu, L.-rong, Zhu, Z., Chan, K. C., Issaq, H. J., Dimitrov, D. S., & Veenstra, T. D. (2007). Improved titanium dioxide enrichment of phosphopeptides from HeLa cells and high confident phosphopeptide identification by cross-validation of MS/MS and MS/MS/MS spectra. *Journal of Proteome Research*, *6*(11), 4150-4162. doi:10.1021/pr070152u

Zhao, H., & Piwnica-Worms, H. (2001). ATR-Mediated Checkpoint Pathways Regulate Phosphorylation and Activation of Human Chk1. *Molecular and Cellular Biology*, *21*(13), 4129-4139. American Society for Microbiology. Retrieved from http://mcb.asm.org/cgi/content/abstract/21/13/4129

Zheng, X. M., Resnick, R. J., & Shalloway, D. (2000). A phosphotyrosine displacement mechanism for activation of Src by PTPalpha. *the The European Molecular Biology Organization Journal*, *19*(5), 964-78. doi:10.1093/emboj/19.5.964

Zhou, Y.-J., Hanson, E. P., Chen, Y.-Q., Magnuson, K., Chen, M., Swann, P. G., Wange, R. L., *et al.* (1997). Distinct tyrosine phosphorylation sites in JAK3 kinase domain positively and negatively regulate its enzymatic activity. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(25), 13850-13855. The National Academy of Sciences of the USA. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9391116

APPENDICES

Supplementary Table: SPASM output. A list of all the identified pockets obtained in the representative library using the canonical RD-pocket as a query and an RMSD cut-off of 1Å.

PDB ID	Kinasefamily	RMSD	Pocket	Enclosed Residue
1ATP	PKA	0.67	K189,H87,R165	T197
1CSN	CK1	0.56	K237,R261,H206	S186
1FMK	SRC	0.94	R155,R175,H201	S158
1FMK	SRC	0.95	R175,R155,H201	Y527
1H1W	PDPK2	0.81	R129, R204, K228	Y126
1JST	CDK2	0.05	R50,R126,R150	T160
1JKS	DAPK	0.58	K45,K46,R48,K86	T87
1O6K	AKT2	0.87	K298,H196,R274	T309
1OL7	Aurora	0.69	R285,R255,R286	S283
1QMZ	CDK2	0	R50,R126,R150	T160
1ZTF	RIO1	0.73	R126, R101, K119	T218
2ERK	ERK2	0.93	R65,R146,R68	T183
2ERK	ERK2	0.27	R68,R146,R170	T183
2H8H	Src1	0.83	R175,R155,H201	Y527
2H8H	Src1	0.88	R175,R155,H201	S158
2I0E	PKC	0.91	K350,K375,K355	T641
2I0E	PKC	0.99	R415,K355,K374	T641
2J0L	FAK1	0.88	R545,K578,R569	S580
2JAM	CAMK1	0.92	R20,R43,K21	T22
20Z0	ZAP70	0.63	R192, R190, R170	T162
2NRU	IRAK4	0.7	K367,R344,K441	S236
2NRY	IRAK4	0.96	R334,R347,R310	T345
2NRY	IRAK4	0.64	K367,R444,K444	S368
2P0C	MER	0.86	H694,R687,R796	S686
2P4I	TEK	0.82	K995,R987,R963	T996
2VX3	DYRK1	0.93	R255,H355,R438	S258
2VX3	DYRK1	0.77	H355,H444,R438	S258

2WO6	DYRK1	0.77	H355,H444,R438	S258
2X7F	TNIK	0.98	R277,H274,K213	S214
2X7F	TNIK	1	R308,K310,K312	T309
2X7F	TNIK	0.99	K311,R313,K312	T309
2X7F	TNIK	0.79	K312,R308,K311	T309
2Y94	AMPK	0.89	R138,K60,R63	T172
2ZVA	LYN	0.66	R400, R366, R390	T398
3A8W	PRKCI	0.82	R318,K258,K277	T555
3ALN	MEK4	0.83	K357,R355,R287	S288
3ANQ	DYRK1A	0.76	R255,H355,R438	S258
3ANR	DYRK1A	0.71	H355,H444,R438	S258
3BLH	CDK9	0.81	R65,R148,R172	T186
3D5W	PLK1	0.93	K194,R161,H214	T196
3D5W	PLK1	0.87	K194,R193,H214	S215
3ENH	KAE1	0.7	R381,R384,K470	Y385
3EQP	CDC4K	0.81	R176, R251, R275	Y284
3HA6	Aurora1	0.8	R180,R255,R286	T288
3HA6	Aurora1	0.67	H306,R286,R255	T288
3HRF	PDK1	0.32	R129,R204,K228	S241
3LJ0	IRE1	0.86	K916,K918,K942	S915
3LJ0	IRE1	0.81	H1044,R1041,K1042	Y1043
3LJ2	IRE1	0.67	R896,K894,R895	T893
3NYX	TYK2	0.7	R1058, 1022, K1046	Y1055
3PFQ	PKC	0.84	R415,K355,K374	T641
3PY3	MAPK14	0.5	K66,R67,R70	T180
3PY3	MAPK14	0.74	R70,R67,R173	T180
3REP	ILK	0.9	H309,K438,R371	S372
3TXO	PRKCH	0.87	K363,K387,K368	T656
3TXO	PRKCH	0.92	R428,K368,K387	T656
3UDB	SRK2	0.85	H231,R232,K228	T229