STRUCTURAL ANALYSIS OF DOMAIN SWAPPING IN THE PROTEIN KINASE

: CRYSTAL STRUCTURE OF HUMAN *STE20* OSR1 KINASE DOMAIN

Approved by supervisory committee

Elizabeth J. Goldsmith,Ph.D.

Diana Tomchick, Ph.D.

Melanie H. Cobb, Ph.D.

Johann Deisenhofer, Ph.D

Stephen R. Sprang, Ph.D

To my family

STRUCTURAL ANALYSIS OF DOMAIN SWAPPING IN THE PROTEIN KINASE

: CRYSTAL STRUCTURE OF HUMAN *STE20* OSR1 KINASE DOMAIN

by

SEUNG-JAE LEE

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

November, 2007

STRUCTURAL ANALYSIS OF DOMAIN SWAPPING IN THE PROTEIN KINASE

: CRYSTAL STRUCTURE OF HUMAN *STE20* OSR1 KINASE DOMAIN

SEUNG-JAE LEE, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2007

Elizabeth J. Goldsmith, Ph.D.

Ste20p (Sterile 20 protein) is a yeast MAP4K involved in the pheromone-responsive MAPK cascade of the mating pathway. Recent studies reveal that its homologs in mammals, *Drosophila melanogaster*, *Caenorhabditis elegans* and other organisms constitute a large emerging group of protein kinases including 28 members in human. The Ste20p family has gained remarkable interest, due to the recent finding that it has various intracellular regulatory effects including the regulation of apoptosis and rearrangement of the cytoskeleton triggering cell-shape changes and cell motility. Moreover, from the viewpoint of structural biology, it is intriguing that the Ste20p family is characterized by the presence of a conserved kinase domain and a noncatalytic region of great diversity at various locations. Despite recent structural studies of Ste20p kinases such as PAK1 or TAO2, the mechanism of kinase regualtion in this family still remains

to be investigated. Thus, in order to deepen our understanding of Ste20p from a structural viewpoint, the OSR1 kinase, which belongs to the GCK VI subfamily, was crystallized. In this dissertation, I present the crystal structure of the OSR1 kinase domain in an inactive conformation. This crystallographic result demonstrates the unexpected finding that the OSR1 kinase domain is dimerized in domain-swapped manner, which is a novel mode of protein-protein interaction in the protein kinase family. The detailed structural analysis shows that αEF helix and the P+1 loop region located in the activation segment were completely swapped between two monomers. The DFG magnesium binding loop and the N-terminal F-helix seem to function as anchors for the hinge loop, based upon the structural alignment with other domain-swapped Ste20p kinases, which were reported after I completed the OSR1 structure determination. Several Ste20p kinases also form domain-swapped dimers in a strikingly similar manner to OSR1.

# Table of Contents

**List of Publications Related to this Dissertation**

1. Lee, SJ, Zhou, TJ, and Goldsmith, E. J. (2006) Crystallization of MAP kinase.

   Methods 40 (3): 224 – 233

2. Lee,SJ, Cobb, M. H., Goldsmith, E. J. (2007) Crystal structure of domain-

   swapped *Ste20* OSR1 kinase domain (to be submitted)

# List of Figures

xiv

# List of Tables

# List of Appendices

# List of Abbreviations

| | |
|---|---|
| **Å** | Ångström |
| **AMPPNP** | Adenosine 5'-($\beta$, $\gamma$-imido) triphosphate |
| **APS** | Advanced Photon Source |
| **ATP** | Adenosine 5'-triphosphate |
| **B factor** | Temperature factor |
| **CCD** | Charge-Coupled Device |
| **CDK** | Cyclin-Dependent Kinase |
| **DAPK** | Death Associated Protein Kinase |
| **DTT** | Dithiothreitol |
| ***E.coli*** | *Eschericheria coli* |
| **EDTA** | Ethylenediamine tetracetic acid |
| **ePK** | eukaryotic protein kinase |
| **ERK** | Extracellular signal-related protein kinase |
| **Fc** | Caculated structure factor amplitude |
| **$F_o$** | Observed structure factor amplitude |
| **GCK** | Germinal Center Kinase |
| **GST** | Glutathione S-Transferase |
| **HEPES** | N-(2-hydroxyethyl)piperzine-N'2-ethanesulphonic acid |
| **I** | Intensity of reflections |
| **IRK** | Insulin Receptor Kinase |
| **ISG** | Interferon Stimulated Gene |

| | |
|---|---|
| **IPTG** | Isopropyl β-D-thiogalactopyranoside |
| **kDa** | Kilo-Dalton |
| **L** | Liter |
| **LB** | Luria-Broth |
| **LOK** | Lymphocyte Oriented Kinase |
| **MAP** | Mitogen Activated Protein |
| **MAPK** | MAP kinase |
| **MAPKK** | MAP kinase kinase |
| **MAPKKK** | MAP kinase kinase kinase |
| **MEK** | MAP kinase/ERK kinase |
| **MR** | Molecular Replacement |
| **Ni-NTA** | $Ni^{2+}$-nitrilotriacetic acid |
| **NKCC1** | $Na^+$-$K^+$-$2Cl^-$ co-transporter-1 |
| **OSR** | Oxidative Stress Responsive |
| **PAK** | p21-activated kinase |
| **PCR** | Polymerase Chain Reaction |
| **PDB** | Protein Date Bank |
| **PEG** | Polyethylene glycol |
| **PKA** | Protein Kinase A |
| **rmsd** | Root mean square deviation |
| **RT** | Rotation Function |
| **SARAH** | Sav/Rassf/Hpo |
| **Sav** | Salvador |

**SDS-PAGE**    Sodium dodecyl sulfate polyacrylamide gel electrophoresis

**SLK**    Sterile20 Like Kinase

**SPAK**    Ste20/Sps1-related proline/alanine-rich kinase

**Ste20p**    Sterile 20 protein

**STK**    Serine /Threonine Kinase

**TAO**    Thousand And One kinase

**TEV**    Tobacco Etch Virus

**Tris**    Tris (hydroxymethyl) aminomethane

**$V_M$**    Matthews number

**WNK**    With No K(Lysine) kinase

# Chapter 1: Introduction

Protein phosphorylation is considered the most common type of post-translational modification utilized in signal transduction machinery. It controls almost all basic cellular process such as growth, differentiation, growth, division, membrane transport, immunity, muscle contraction, organelle trafficking, metabolism, learning and memory (Manning, Plowman et al. 2002; Manning, Whyte et al. 2002). The eukaryotic protein phosphorylation event is conducted by a single superfamily of eukaryotic protein kinases (ePKs) that share a highly conserved catalytic domain in terms of their sequences and overall structure. Several atypical protein kinases, however, are shown to contain low structural similarity to ePKs. Protein kinases catalyse the transfer of the γ-phosphate from ATP to specific residues in protein substrates, which are mainly Ser, Thr and Tyr.

Given the unique significance in signal transduction, it is no wonder that protein kinase superfamily is one of the largest families of genes in eukaryotes, accounting for more than 2 % of the whole genome (Zhu 2000; Caenepeel, Charydczak et al. 2004). Moreover, it is considered that about 30 % of all cellular proteins are post-translationally modified by phosphorylation on at least one residue (Pinna and Ruzzene 1996; Cohen 2000). The current understanding of protein phosphorylation conducted by kinases suggests that protein kinases differ remarkably in the number of substrate phosphorylation sites they actually phosphorylate. For instance, the mammalian protein kinase MEK1 (mitogen-activated protein kinase (MAPK) and extracellular signal-regulated kinase (Willoughby, Perkins et al.) kinase-1) phosphorylates only four sites, two in ERK1 and two in ERK2, whereas calcium-calmodulin (Gupta, Campbell et al.)

1

kinases and cyclin-dependent kinases (CDKs) phosphorylate more than hundreds of sites. In vitro systematic proteomic studies in Saccharomyces cerevisiae imply this possibility: in one study, kinases phophorylated between 1 and 256 substrates, and in a second study, cyclin-dependent kinase-1 (Cdk1) from yeast phosphorylated hundreds of substrates (Ubersax 2003; Ptacek 2005).

The spatial and temporal regulation of protein phosphorylation on therine/serine or tyrosine is demonstrated to be crucial in cellular growth and development. Now, it is clear that the complex mechanism of protein kinase regulation is responsible for the delicate substrate recognition system. Thus, the abnormal deregulation of kinase activity in the wrong place and wrong time turns out to have detrimental consequences, giving rise to cell transformation and cancer. It is no wonder the first oncogene to be discovered, *v-src*, encodes an aberrantly regulated tyrosine kinase (Levinson, Oppermann et al. 1978; Eckhart, Hutchinson et al. 1979).

In terms of sequence and structure, the serine/threonine and tyrosine kinases form a highly conserved superfamily that has distinct functional and structural features from the histidine kinases and other phosphotransfer enzymes.

Whereas the several hundred eukaryotic protein kinases are characterized by the same catalytic scaffold, a variety of regulatory mechanisms have existed that confer individual protein kinases an ability to function downstream of the specialized input signals that turn them on. Protein kinases are usually kept off and the activation mechanism of catalytic activity is often under a set of layers of delicate control, ranging from the binding of allosteric effectors to subcellular localization of the given enzyme.

Protein kinases function as molecular switches that can adopt two distinct states via conformational changes: an "on" state that fully active, and an "off" state that has minimal activity. Basically speaking, it can be said that all protein kinases catalyze the identical chemical reaction, which involves the transfer of the γ-phosphate of ATP to hydroxyl group of serine, threonine or tyrosine.

It is clearly shown from the structural studies that all the protein kinases adopt very similar conformation upon activation (Knighton, Zheng et al. 1991; Yamaguchi and Hendrickson 1996; Canagarajah, Khokhlatchev et al. 1997). However, protein kinases have evolved to have structurally distinct off-state conformation, which is not required to satisfy the chemical restraints for active kinase activity.

**The Structure of the Kinase Domain**

The structure of a eukaryotic protein kinase domain was first presented by X-ray crystallography in the case of cyclic AMP-dependent protein kinase, a Ser/Thr kinase also known as protein kinase A (PKA) (Knighton, Zheng et al. 1991). This crystal structure of PKA complexed with Mn-ATP and the substrate mimic PKI affords the structural snapshot for protein kinase in an active conformation, which is prime and ready for phosphotransfer (**Figure 1-1**) (Knighton, Zheng et al. 1991).

The protein kinase fold, which is extremely well conserved among serine/threonine and tyrosine protein kinases, is divided into two lobes. The smaller N-terminal lobe consists of five-stranded β sheet and one α helix, called helix αC. The C lobe is larger and heavily helical. ATP is bound in a cleft between the two lobal substructures of protein kinase and sits beneath a highly conserved loop linking β1 and

β2 strands. This loop is termed phosphate binding loop or P-loop, which consists of a conserved glycine-rich sequence motif (GXGXΦG) in which Φ contain aromatic side chain such as tyrosine or phenylalanine aromatic side chain. In the absence of ATP, this P-loop becomes very flexible owing to the presence of the glycine residue, which explains why the binding of small molecule is favorable.



**Figure 1-1. Ribbon diagram of the catalytic subunit of PKA representing the structure of the protein kinase core (PDB code: 1ATP).**

It is shown from the previous observation that large conformational disruptions were induced by some of small-molecule inhibitors in the loop through the interaction with the conserved aromatic residue (Mohammadi, McMahon et al. 1997; Schindler, Bornmann et al. 2000).

In order for substrate to approach and bind active site, a centrally located loop known as the activation loop should be in an extended conformation across the front end of the nucleotide binding pocket, close to the γ-phosphate of ATP. The physical platform for substrate binding is provided by this 20~30 residue-length activation loop. Usually when protein kinase is active, this loop is phosphorylated on the specific residues including serine, threonine or tyrosine. This activation loop is stabilized in an open and extended conformation suitable for substrate binding and subsequent catalytic reaction via phosphorylation event, which is usually conducted by other activating protein kinase or autophosphorylation.

The precise physical arrangement of several catalytic residues must be satisfied for optimal subsequent phosphotransfer event. It is well known that those catalytic residues are very well conserved among all known kinases. For example, Asp166 and Asn171 (PKA numbering) extrude from a highly conserved loop at the base of the active site called the catalytic loop. Asp166 is met by the attacking hydroxyl side chain of the substrate, when Asn171 involves in hydrogen bonding that ensure the proper position of Asp166. In addition, Asp184, absolutely conserved catalytic residue, and Asn171 are required for the binding of two divalent cations such as Mn or Mg, which engage in nucleotide recognition. This Asp184 comprises part of the highly conserved Asp-Phe-

Gly (DFG) motif located at the base of acticvation loop.  The spatial orientation of this DFG structure is directly linked to phosphorylation event of activation loop.

In the N-terminal lobe, Lys72 makes critical interaction with the α- and β-phosphate groups, orienting them suitable for enzymatic reaction.  It is clearly shown that Lys72 is located deep between the interlobal cleft, in which it makes crucial contact with Glu91 from the αC helix via an ionic interaction.

When the protein kinase is activated or inactivated, the activation loop undergoes noticeable conformational changes (Johnson, Noble et al. 1996).  As indicated in **Figure 1-2**, insulin receptor kinase (IRK) is activated by phosphorylation of three tyrosine residues in the activation loop.  In the case of unphosphorylated state, the activation loop actually folds back onto the base of the loop where active site is located, preventing the binding of both nucleotide and peptide substrate (Hubbard, Wei et al. 1994).  Upon phosphorylation, the distant part of activation loop is moved away from the active site and undergoes a dramatic conformational change, which makes protein kinase permissive for substrate binding and subsequent catalysis (Hubbard 1997).

**Active**
**Insulin Receptor Kinase**

**Iactive**
**Insulin Receptor Kinase**
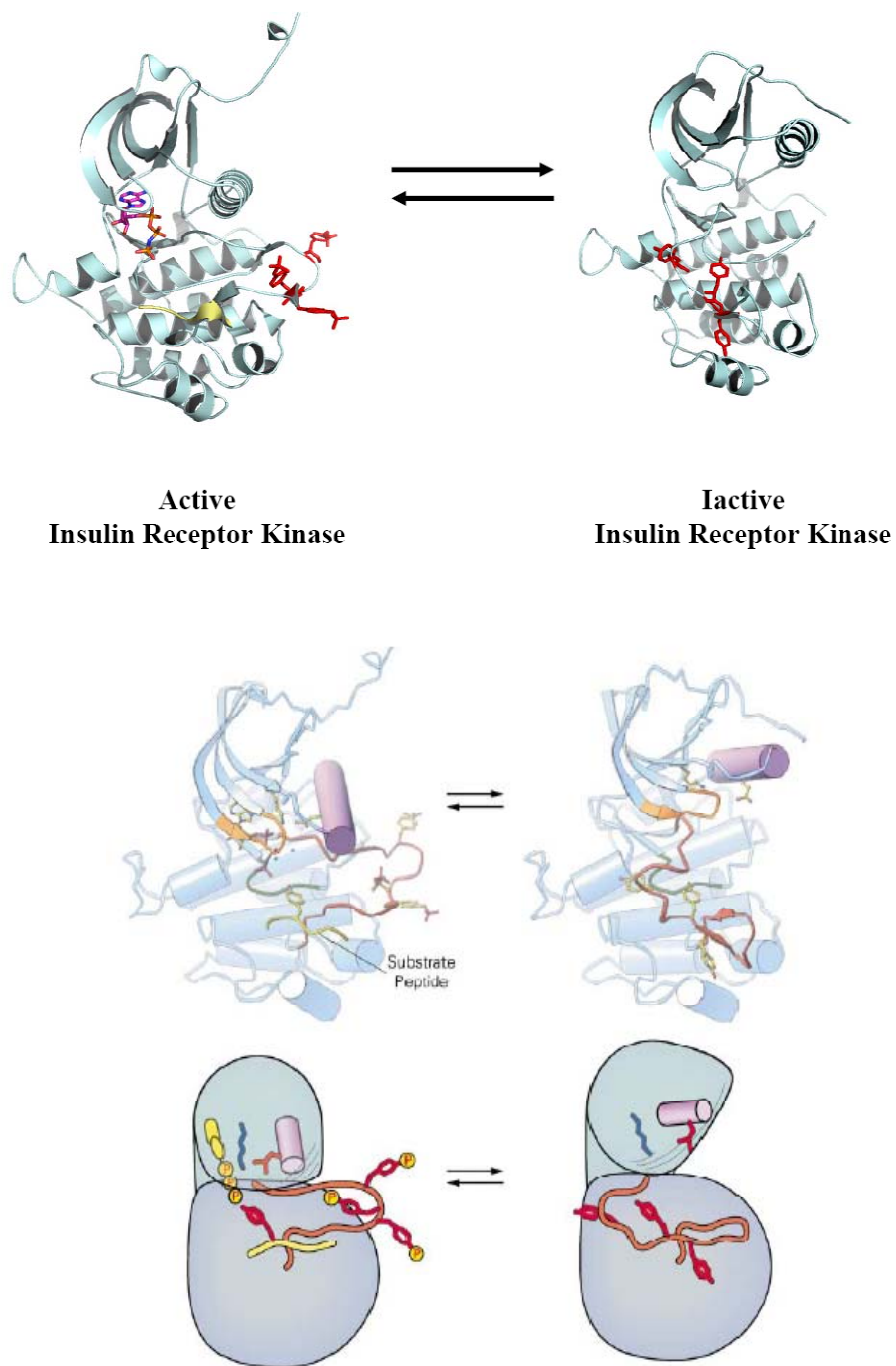
**Figure 1-2. Ribbon diagram and schematic diagram representing the structures of insulin receptor kinase** (IRK, a tyrosine kinase) (TOP) The crystal structures of active (phosphorylated) and inactive (unphosphorylated) insulin receptor kinase. The

phosphorylatable residues in the activation loop are colored in red. (Bottom) The schematic diagram characterizes the conformational changes involving in the regulation of kinase activity, focusing on the αC helix and activation loop. The catalytic lysine and glutamate are colored in each diagram. The diagrams on the bottom were adapted from (Huse and Kuriyan 2002).

The number of phosphorylation site as well as its location within the activation loop varies among the individual kinases. Usually, one phosphate group forms an ion pair interaction with a conserved arginine (Arg165 in PKA numbering) situated at the base of the catalytic loop. This ion pair between arginine from catalytic loop and phosphorylated residue from activation loop seems to function as mediator involved in rotating and orienting the DFG motif into proper position for catalysis (Hubbard, Mohammadi et al. 1998).

In some cases, the conformational change triggered by the phosphate group from activation loop is coupled to subsequent local rearrangement of neighboring resides and gives rise to a set of interactions that propagate well into the N terminal lobe in a more prominent fashion. That's the case in ERK2. The mitogen-activated protein (MAP) kinase ERK2 is phosphorylated at a threonine and a tyrosine residue within activation loop. Then, phosphorylated threonine pThr183 now plays a crucial role as mediator by making contacts with αC helix, resulting in the closure of the N and C lobes such that ERK2 becomes active (Canagarajah, Khokhlatchev et al. 1997). Furthermore, this set of conformational changes initiated by the phosphate group within the activation loop is coupled to the oligomerization status of ERK2. A C-terminal extension of ERK2 binds the surface of the N terminal lobe (Zhang, Strand et al. 1994). Following the phosphorylation at threonine 183 in ERK2, a network of conformational changes near C-

terminal extension is developed to help expose a hydrophobic surface, which drives the formation of homodimers **(Figure 1-3)** (Canagarajah, Khokhlatchev et al. 1997). Also, it is shown that the nuclear localization of ERK2 requires the formation of homodimers via hydrophobic interaction (Khokhlatchev, Canagarajah et al. 1998).

More interestingly, the activation loop of each protein kinase undergoes a unique conformational change when the kinase is off. This fact actually promoted the practical idea that each inactive form of protein kinase can be superior drug target in terms of its specificity to active one, which constrains a rigid orientation of key catalytic residues resulting in very similar conformation. One of such examples is the anticancer drug Gleevec, an efficient therapeutic for the treatment of chronic myelogenous leukemia. As explained above, this drug binds to an inactive Abl kinase specifically, promoting the stabilization of activation loop in certain conformation that mimics bound substrate (Druker, Talpaz et al. 2001).

Inactive
MAP Kinase (ERK2)

Active
MAP Kinase (ERK2)

Dimerization
Interface

Inactive
MAP Kinase
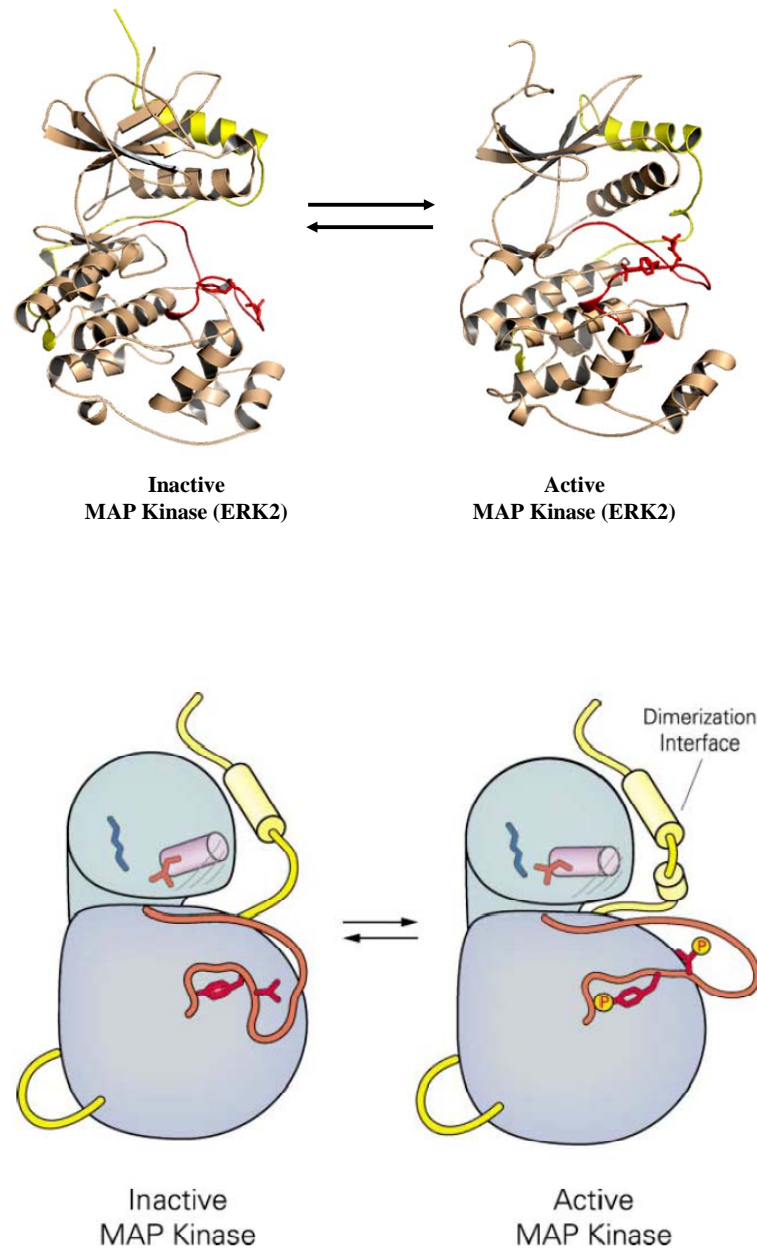
Active
MAP Kinase

**Figure 1-3**. The Regulation of MAP kinase via phosphorylation of the activation loop. The bottom diagram was adapted from (Huse and Kuriyan 2002)

**The *Ste20* group kinases**

The emerging group of kinases related to the budding yeast Ste20p (sterile 20 protein) kinases has recently received noticeable attention and also been studied extensively, mainly due to its functional novelty and role as regulators of MAP kinase cascades. Currently, there are approximately 30 Ste20-related kinases identified in mammals, as well as homologs in *Drosophila*, *C elegans* and other organisms (Sells and Chernoff 1997; Kyriakis 1999).

This *Ste20* group kinase is classified into two subgroups, which include germinal center kinases (GCKs) and p21-activated kinases (PAKs), based upon its structure and regulation. The PAKs comprise the six enzymes nearer in characteristics such as sequence homology to Ste20p itself. It is shown that each PAK contains a C-terminal catalytic domain and N-terminal regulatory domain with a small G-protein binding motif. PAKs have been shown activate MAPKs (primarily JNK and p38) as well as to affect disassembly of the actin cytoskeleton and apoptosis. On the other hand, the GCKs are distinct from PAKs in that they have N-terminal catalytic domains (Sells and Chernoff 1997; Kyriakis 1999).

In general, Ste20 kinases are shown to be involved in a variety of intracellular regulatory effects including the rearrangement of the cytoskeleton generating cell-shape change, cell motility and the regulation of apoptosis (Sells and Chernoff 1997; Kyriakis 1999).

More interestingly, recent studies show that the mitogen-activated protein kinase (MAPK) cascades are activated by most Ste20 group kinases. It is well known that the MAPK cascades are essential in a wide range of cellular events, transmitting extracellular

stimuli including growth factors, environmental stresses and cytokines to activate transcription factors, involving in the regulation of a set of gene expression (Ip and Davis 1998; Widmann, Gibson et al. 1999).

*Ste20p as a MAP4K in yeast*

Ste20p was originally identified and characterized as a key kinase in the mating pathway of budding yeast *Saccharomyces cerevisiae* (Wu, Whiteway et al. 1995). This pathway is activated by the binding of a short peptide pheromone to a pheromone receptor, which then generates the activation of heterotrimeric G protein and gives rise to the activation of the yeast MAPK pathway formed by Ste11p (MAP3K), Ste7p (MAP2K) and Fus3p and Kss1p (MAPK) (Gustin, Albertyn et al. 1998). Especially, the recent study demonstrating the direct phosphorylation of Ste11p by Ste20p indicates that Ste20p can function as an MAP4K (Wu, Whiteway et al. 1995; Drogen, O'Rourke et al. 2000). Even though epistatic study suggests Ste20p as direct upstream element of Ste11p (Wu, Whiteway et al. 1995; Drogen, O'Rourke et al. 2000), this does not indicate that *Ste20p* is located in the linear sequential phosphorylation cascade module. Rather, it should be understood that the two different signaling modules meet on the plasma membrane after activation by a pheromone and relay its signal via phosphorylation event. Thus, Ste20p should be regarded as one of the upstream activators of the MAPK pathways.

*PAKs and GCKs as the mammalian homologs of Ste20p*

In mammals, Ste20p-related protein kinases are divided into two subfamilies including the PAK and the GCK families (Sells and Chernoff 1997; Kyriakis 1999).

They are characterized by the relative location of their kinase domain to non-catalytic regulatory domains. The kinase domains of PAKs are located at their C-termini, whereas those of GCKs are at their N-termini. It is shown that the structures of the kinase domain are conserved among PAKs, GCKs and Ste20p. Interestingly, in the subdomain VIII of Ste20 group kinases, there is a distinct conserved sequence, GTPyWMAPEv (small letters indicate less conservation), termed the Ste20 signature sequence (Sells and Chernoff 1997). Due to the fact that the subdomain VIII is mainly responsible for the recognition of substrates in a variety of protein kinases, it is suggested that the Ste20 group kinases might confer highly similar substrate specificities. It thus follows that the mammalian Ste20 group kinases may activate MAP3Ks in the same way that the yeast MAP3K.

*Human Ste20 group kinase*

Thus far six human PAK and 22 human GCK family members have been cloned and identified (Dan, Watanabe et al. 2001). Currently, the number of Ste20 group kinases is likely to be in the neighborhood of 30 in humans. Based upon the kinase domain conservation, these human Ste20 kinases are classified into subfamilies (Dan, Watanabe et al. 2001). The PAK and GCK families are divided into two and eight distinct subfamilies, respectively. Given the phylogenetic analysis based upon the relative location of catalytic and non-catalytic region, it is strongly suggested that all the kinase in each *Ste20p* subgroup should have their distinct structural features associated with functional regulation as shown in **Figure 1-4**.
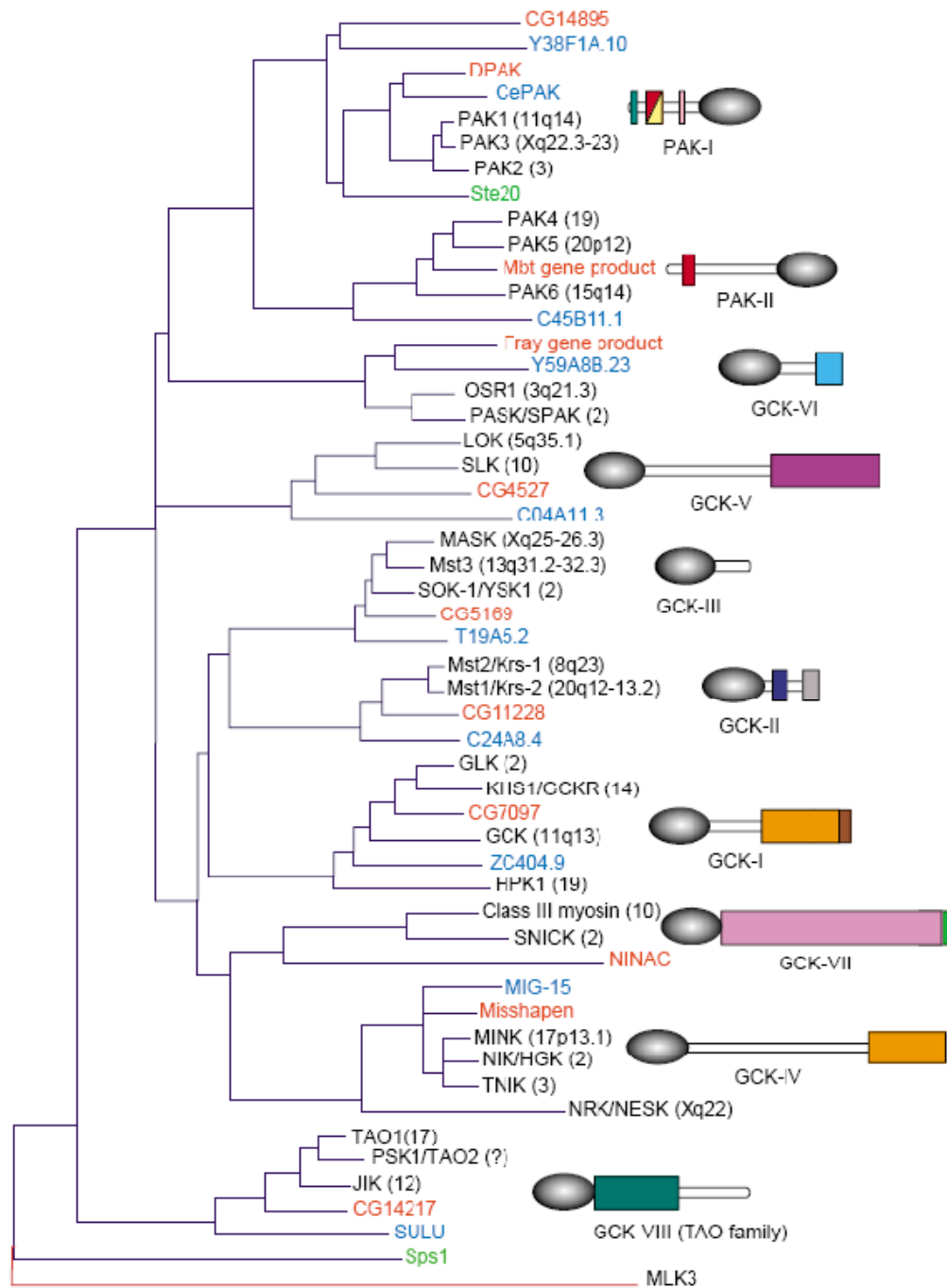
**Figure 1-4. Phylogenetic relations among mammalian Ste20 group kinases and their schematic structures in each subfamily**. The protein sequence of the kinase domain of every *Ste20* group kinase was from human (Black), *Drosophila* (red) and *Caenorhabditis elegans* (blue) databases. The budding yeast *Ste20p* and *Sps1p* are incorporated with the phylogenetic analysis to provide references (green). The black ovals represent kinase domains. The rectangles with various colors represent conserved regulatory domains and the white bars indicate less conserved, variant regions. This figure was adapted from (Dan, Watanabe et al. 2001).

*Ste20 group kinases in apoptosis, morphogenesis and cytoskeletal rearrangement*

Many of the Ste20 homologs are utilized in signal transduction, including apoptosis, morphogenesis and cytoskeletal rearrangement events such as cell motility and cell-shape changes (Dan, Watanabe et al. 2001). Some Ste20 group kinases are associated with inducing or preventing apoptosis, whereas others are involved in regulating cellular event induced by apoptosis.

For example, it is well known that activation by Fas by ligand binding induces a sequential cleavage of caspases in a typical apoptotic pathway. The activated caspases then cleave the most downstream effector pro-caspases to generate activated effector caspases. In the following event, these activated effector caspases cleave and release various Ste20 group kinases such as Mst1, Mst2, HPK1, SLK and PAK2 (Lee, Murakawa et al. 1998; Chen, Meyer et al. 1999; Sabourin, Seale et al. 2000; Dan, Watanabe et al. 2001). Some GCKs, subfamily members of Ste20p, such as HPK1 and SLK are activated during apoptosis (Sabourin and Rudnicki 1999). It is also suggested that PAK-II subfamily kinases are involved in the reorganization of actin cytoskeleton, together with PAK-I (Sells and Chernoff 1997). It is reported that PAK4 can induce filopodia formation through the rearrangement of the cytoskeleton, even though its mechanism is

not fully understood (Abo, Qu et al. 1998). In addition, it is suggested that PAK-II subfamily members might play a role in neuronal morphogenesis as a deficiency of PAK-II in Drosophila results in severe defects in central brain structure (Melzig, Rein et al. 1998).

**_Ste20_ OSR1 Kinase**

The GCK (Germinal Center Kinase) VI subfamily of _Ste20p_ comprises two kinases: SPAK (<u>S</u>te20-related <u>p</u>roline <u>a</u>lanine-rich <u>k</u>inase) and OSR1 (<u>O</u>xidative <u>s</u>tress <u>r</u>esponse <u>1</u>). These two closely related kinases share about 65% sequence identity and contain well-conserved C-terminal regions termed PF1 and PF2 domains. This unique region was demonstrated to be a critical domain of the kinase, interacting with RFxV motifs in membrane transport systems such as cation-chloride cotransporters and chloride channel but also associating with a variety of proteins including heatshock protein 105, WNK4 (<u>W</u>ith <u>N</u>o lysine <u>K</u>inase), apoptosis-associated tyrosine kinase, gelsolin, the actin cytoskeleton, and p38 MAPK (Johnston, Naselli et al. 2000; Piechotta, Garbarini et al. 2003; Denton, Nehrke et al. 2005).

It is well established that SPAK and OSR1 play integral roles in cellular signal transduction such as cell-cycle control, growth and response to stress (Strange, Denton et al. 2006). Recent studies clearly show that this family is intimately involved in cell volume and [Cl-] regulation. SPAK and OSR1 are closely related members of a germinal centre kinase subgroup (GCK-VI). OSR1, the product of a more ancestral gene, is expressed in most tissues, whereas SPAK, which probably evolved after the amphibia, is

found mainly in neurons and transporting epithelia, with its expression closely paralleling that of NKCC1 ($Na^+/K^+/2Cl^-$ (Delpire and Gagnon 2006).

Although there was strong evidence for the binding of SPAK and OSR1 to the N-terminus of NKCC1 and NKCC2 and a binding motif was identified (Piechotta, Lu et al. 2002), it is yet to be decided whether the kinases could phosphorylate the transporters, so it is believed that SPAK and OSR1 could act as scaffolding molecules, bringing together components of a stress-response complex on the surface of the cotransporters (Piechotta, Garbarini et al. 2003).

Expression studies also show that NKCC1 can be phosphorylated and activated by OSR1 as long as this was co-expressed with another kinase, WNK4, which had been earlier identified as a potential binding partner in yeast two-hybrid screens (Piechotta, Garbarini et al. 2003). There is now clear evidence that WNK4, and the related WNK1, phosphorylate and activate both OSR1 and SPAK.

**Description of dissertation research**

This dissertation presents the determination and detailed analysis of the crystal structure of dimerized domain-swapped human *Ste20p* OSR1 kinase domain. This crystal structure provides unique insights into the structural basis for domain swapping in the *Ste20p* protein kinase together with other "swapped" protein kinase structures. Prior to the detailed structural analysis of domain-swapped OSR1 kinase structure, various efforts to obtain soluble construct as well as diffraction-quality protein crystal are discussed extensively. In addition to expression, purification and crystallization of this target protein OSR1, the molecular replacement method I chose to utilize in order to solve this crystal structure was also described in detail.

In summary, the research work described in the following chapters (Chapters 2 to 5) covers two major themes: (1) Expression, Purification and Determination of Crystal Structure of Domain-Swapped *Ste20p* OSR1 Kinase Domain; (2) Detailed Structural Analysis of observed Domain-Swapping by comparison of OSR1 with other "swapped" protein kinase structures.

Chapter 2 describes the expression, purification and crystallization of human *Ste20p* OSR1 kinase.

Chapter 3 covers the structure determination of human *Ste20p* OSR1 kinase domain, focusing on the twilight-zone molecular replacement method utilized to address the problem of taking MR approach for protein kinase structures. In addition to novel MR approach for protein kinases this chapter 3 involves model building and refinement of crystal structure.

Chapter 4 presents the structural analysis on the crystal structure of human *Ste20p* OSR1 kinase domain. Above all, this chapter 4 deals with the sequence alignment analysis for understanding the origin of this intriguing domain-swapping molecular event. Also, the relationship between sequence and domain swapping in the protein kinase is explored.

Chapter 5 addresses my conclusion on this structural analysis of domain swapping in the Ste20p protein kinase and discusses the future directions for research towards understanding the underlying physico-chemical mechanism of domain-swapping and its fuctional role in the signal transduction pathway.

Chapter 6 describes the expression, purification of human Aurora B kinase and Incenp.

Chapter 7 describes the expression, purification of *Ste20p Hippo* kinase and *Sav*. It also covers the extensive crystallization trials for *hippo* c-terminal SARAH domain.

Chapter 8 describes the expression, purification of ISG15 and the following crystallization trials.

# Chapter 2: Expression, Purification and Crystallization of human *Ste20* OSR1 kinase

## Introduction

In order to obtain a high-quality diffracting protein crystal of a given target protein, a set of problems should be addressed related to its biochemical, biophysical properties. Whereas *in vivo* or *vitro* biochemical experiments usually dealing with a small scale of protein amount less than milligram range, crystallization methods for solving a crystal structure requires a high level of expression as well as solubility and stability of a given target protein.

Like other interesting proteins that have been crystallized and solved, the human *Ste20p* OSR1 kinase also presents its unique challenge in terms of expression, purification and crystallization followed by its subsequent improvement of crystal diffraction quality.

To address such challenges related to expression, purification and crystallization of my target protein, prokaryotic and eukaryotic expression systems were extensively utilized with a series of various constructs and also optimized for this human *Ste20p* OSR1 kinase.

This chapter describes the unique efforts to produce a high-quality diffracting OSR1 kinase crystal via a series of experimental procedures. First, I will briefly describe all the experimental procedures for construct cloning, expression, purification and

crystallization throughout the [Methods and Procedures] section.  Secondly, I will discuss all the utilized strategies optimized for human *Ste20p* OSR1 kinase in greater detail through the next section [Results and Discussion].

# Materials and Procedures

**Cloning of Constructs**

Each construct was subcloned into the pHis-parallel vector containing N-terminal 6X histidine-tag.  All the constructs were generated based upon previous sequence-domain analysis.  The PCR reaction condition I used is as follows.

2 μl     template

1 μl     primer 1 (50pmol/μl)

1 μl     primer 2 (50pmol/μl)

5 μl     *Pfu* reaction buffer (10X)

1 μl     *Pfu* enzyme

1 μl     dNTP (10mM) mixture

39 μl    water_____

50 μl    total reaction volume

This mixture was applied to PCR reaction to amplify the target sequence.

1st Cycle:            94 °C for 2 minutes to denature the template

2nd ~ 30th Cycle:    94 °C for 30 seconds to denature the template

                     58 °C for 40 seconds for primer annealing

72 °C for 90 seconds for primer extension

After PCR reaction, the PCR product was checked by DNA electrophoresis and purified. And this PCR product was digested by restriction enzymes as follows and applied to ligation reaction by T4 DNA ligase, followed by DNA sequencing.

30 µl PCR (purified)

2 µl restriction enzyme I

2 µl restriction enzyme II

0.5 µl BSA (100X)

5 µl restriction enzyme buffer

10.5 µl water

50 µl  Total reaction volume

This mixture was applied to digestion reaction at 37 °C for 2 hrs, followed by subsequent purification for ligation procedure.

**Expression of human *Ste20* OSR1 kinase in *E.coli***

A. Transformation

    1. Mix 20 µl of Rosetta2(DE3) cells with 1 µl of DNA plasmid vector and incubate the mixture on ice for 5 minutes

    2. Heat-shock was applied to cell-DNA mixture for 40 sec by using 42°C water bath.

    3. Transfer the cell-DNA sample onto ice promptly and keep it on ice for 3 minutes.

4.  Add 80 μl of sterilized LB or SOC medium and grow the cells for 1 hour at 37°C

     in the temperature-controlled shaker

5.  Apply the grown cells onto the LB plate containing


B.  Cell Growth

   1.  Streak out *E. coli* strain Rosetta2 (DE3) transformed with the pHis-pararell

        vector  encoding human OSR1 kinase domain on a LB plate containing 100

        μg/ml of ampicillin (or 50 μg/ml of carbenicillin).

   2.  Incubate the LB plate overnight at 37°C to grow the transformed colonies.

   3.  Inoculate 100 ml of LB medium containing 100 μg/ml of ampicillin in a

        250 ml flask with a single colony, and grow the cells at 37°C in a shaker

        (250~300 rpm) overnight.

   4.  Use 10 ml of the overnight-grown culture to inoculate 1L of sterilzed LB

        medium supplemented with 3% ethanol and 100 μg/ml of ampicillin.

   5.  Grow the cells at 37°C until OD at 595 nm reaches about 0.8. (It takes

        about 8 ~ 9 hours).

   6.  Add 0.5 ~ 1 mM IPTG to induce OSR1 expression and grow the cells at

        30°C for overnight

   7.  Harvest the cells by using a Beckman J6-MI centrifuge with a 6x1 L

        swing-out rotor for 30 min at 4000 rpm (at 4°C).

   8.  Resuspend the cell pellet with about 35 ml of chilled Ni-binding buffer and

        treat it with the protease inhibitor cocktail.

   9.  Flash-freeze the resuspended cell pellet in liquid nitrogen and store in –

80°C freezer.

**Expression of human *Ste20* OSR1 kinase in Insect cell**

The overall scheme of baculovirus protein expression system is presented in **Figure 2-1**. The first step for producing target protein in insect cell (SF9) utilizing baculovirus is to clone the gene of interest into pFastBac donor plasmid by using standard cloning procedure. Once the gene fragment was cloned into pFastBac donor plasmid, this DNA was transformed into DH10Bac for transposition into bacmid.

A. Transposition

   1. Prepare Luria Agar plates containing:

      50 µg/ml kanamycin

      7 µg/ml gentamycin

      10 µg/ml tetracycline

      100 µg/ml Bluo-gal

      40 µg/ml IPTG

   2. Thaw DH10Bac competent cells on ice.

   3. Dispense 100 µl of the cells into 15 ml tube.

   4. Add recombinant donor plasmid (5 µl) and gently mix this DNA into cells by tapping the tube.

   5. Incubate the mixture on ice for 30 min.

   6. Heat shock the DNA-cell mixture by transferring to 42 °C water bath for 45 sec.

7. Put the mixture on ice for 3 min.

8. Place the mixture in a shaking incubator at 37 °C with 225 rpm for 4 hrs.

9. Make the serial dilution to cells ($10^{-1}$, $10^{-2}$, $10^{-3}$).

10. Place 100 μl of each dilution on the plates and spread it over the plate.

11. Incubate the plates for 24 to 48 hrs at 37 °C (Since colonies are very small, grow until blue colonies appear discernibly blue for about more than 24 hrs)

B. Recombinant Bacmid DNA Isolation

1. White colonies are selected from a plate with approximately 100 to 200 colonies.

2. After picking up 10 white candidate colonies, restreak to fresh plates to confirm the white colony phenotype. Incubate overnight at 37 °C.

3. Set up a liquid culture with antibiotics for isolation of recombinant bacmid from a single white colony

4. Detailed experimental procedure for isolating large plasmid like bacmid DNA (>100kb) used for this experiment was explained in [Bac-to-Bac Baculovirus Expression Systems] (Invitrogen).

5. After isolation, store bacmid DNA at -20 °C. Please avoid repeated freeze/thaw cycles in order to prevent reduction in transfection efficiency.

C. Transfection of Sf9 cells with Recombinant Bacmid DNA

1. Seed $9 \times 10^5$ cells per 35-mm well (of 6 well-plate) in 2 ml of Sf-900 II SFM. Always use cells only from 3~4 day old suspension culture in mid-

log phase.

2. Allow cells to attach at 27 °C for at least 1 hrs.

3. Prepare the following solutions in 12 x 75-mm sterile tubes:

   Solution A: For each transfection, dilutes ~ 5 μl of mini-prep bacmid DNA
   into 100 Sf-900 II SFM without antibiotics.

   Solution B: For each transfection, dilute ~ 6 μl CellFECTIN reagent into
   100 Sf-900 II SFM without antibiotics.

4. Combine the solution A and B, mix gently, and incubate for 15 to 45
   minutes at room temperature

5. Wash the cells with 2 ml of Sf-900 II SFM without antibiotics.

6. For each transfection, add 0.8 ml of Sf-900 II SFM to each tube containing
   lipid-DNA complexes.  Mix very gently.  Aspirate wash media from cells
   and overlay the diluted lipid-DNA complexes onto cells.

7. Incubate the cells for 5 hrs in a 27 °C incubator

8. Remove the transfection mixtures and add 2 ml of Sf-900 II SFM without
   using antibiotics.  Incubate cells in a 27 °C incubator for 72 hrs

9. Harvest virus from cell culture medium at 72 hrs post-transfection.


D.  Harvest and Storage of Recombinant Baculovirus

1. To harvest virus from the transfection, transfer the supernatant (2 ml) to a
   capped, sterile tube.  Spin down for 5 min at 500 x $g$ and transfer the virus-
   containing supernatant to a fresh tube.

2. Store the virus at 4 °C protected from light.

E.   Infection of Insect Cells with Recombinant Baculovirus Particles

1.  To amplify the virus, about 1:50 dilution of virus was used with Sf-900 II

    SFM without antibiotics.  Initial 2 ml of supernatant was termed as P0.  The

    next amplified virus was designated as P1.  The virus was harvested usually

    at 48 h post-infection

2.  For protein expression, usually P4 or P5 was used.

    (For the best result, cells were harvested at different time intervals: 24h,

    48h, 72h and 96 h)

3.  Harvest the cells by using a Beckman J6-MI centrifuge with a 6x1 L

    swing-out rotor for 30 min at 4000 rpm (at 4°C).

4.  Resuspend the cell pellet with about 35 ml of chilled Ni-binding buffer and

    treat it with the protease inhibitor cocktail.
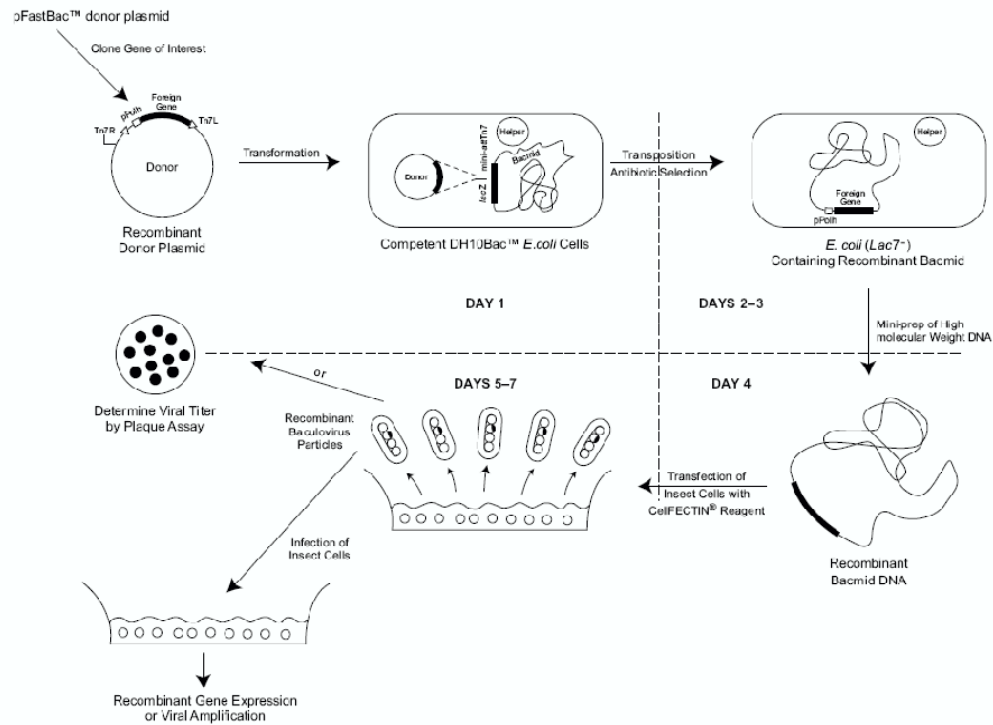
**Figure 2-1. Generation of recombinant baculoviruses and gene expression with insect cell expression system.** This figure was adapted from instruction manual for [Bac-to-Bac Baculovirus Expression Systems] (Invitrogen)

**Purification of human *Ste20* OSR1 kinase**

All the main buffer components used for purification step was described in **Table 2-1**.

A.  Preparation of the Cell-free Extract (from *E.coli*)

1.  Thaw the frozen cells (taken from freezer-storage) at 4°C cold room or on ice.

2.  Following the treatment of cell lysate with lysozyme and bezonase nucleases at 4°C for about 1 hr, sonicator was used to break the thawed cells on ice for about 3 minutes (10 sec on and 5 sec off)

3.  By utilizing centrifugation of the cell lysate at 35000 rpm for 1 hr (4°C), the spinned-down cell debris was separated from the supernatant cell extract containing the target protein.

4.  The cell extract cleared by 0.45 μm syringe filter was transferred to the prechilled beaker for the next purification step.

A-1.  Preparation of the Cell-free Extract (from insect cells)

1.  Thaw the frozen cells (taken from freezer-storage)

2.  Tissue homogenizer was applied to break the insect cells (about 25 ~ 30 times of piston movement)

3.  By utilizing centrifugation of the cell lysate at 35000 rpm for 1 hr (4°C), the spinned-down cell debris was separated from the supernatant cell extract containing the target protein.

4. The cell extract cleared by 0.45 μm syringe filter was transferred to the prechilled beaker for the next purification step.

B. Immobilized Metal Affinity Chromatography

1. By using a peristaltic pump (Amersham Biociences), the cleared cell extract was applied to a fast-chelating sepharose column (Amersham Pharmacia) loaded with 0.1 mM NiSO and pre-equilibrated with 5~10 column volume of Ni-binding buffer.

2. In order to remove non-specific bound proteins, after loading the sample (the cleared cell extract), the column was washed by 3~4 column volume of washing buffer with the increasing concentration of imidazole (from 10 to 40 mM) using step gradient.

3. Elute the target protein with 250 mM imidazole-containing Ni-elution buffer and pool the eluted fractions.

C. TEV-protease Digestion

1. The eluted sample fractions from Ni-column was promptly was mixed with the MonoQ A buffer and transferred into dialysis bag for buffer change at 4°C for overnight.

2. Additionally, about 1/30 ~ 1/50 ratio of TEV-protease was added for tev-protease digestion for removing N-terminal 6X His tag.

D. Anion Exchange Chromatography

1. Spin down the dialyzed protein solution to remove the precipitated debris using 4000 rpm centrifugation.

2. Clear the dialyzed protein solution by 0.22 μm filter syringe

3. Load the filtered protein sample onto a MonoQ anion-exchange column (Pharmacia).

4. Following the washing step with the 3 column-volume of MonoQ A buffer, the protein sample was applied to step gradient for elution of target protein sample.


E. Size-Exclusion Chromatography

1. The eluted protein sample from anion-exchange chromatography was pooled and concentrated up to the total volume of 2~3 ml.

2. Following the filtration of sample protein by 0.22 μm filter syringe, the filtrate was loaded onto Superdex-75 or 200 size-exclusion column. The eluted sample was pooled and collected. The protein sample for crystallization was usually concentrated up to ~ 12 mg/ml.

**Table 2-1 Summary of reagents used for purification of OSR1 kinase domain**

| | |
|---|---|
| Ni-binding buffer | 50 mM Tris HCl, pH 8.0 |
| | 300 mM NaCl |
| | 10 mM Imidazole |
| Ni-elution buffer | 50 mM Tris HCl, pH 8.0 |
| | 300 mM NaCl |
| | 250 mM Imidazole |
| Mono Q buffer A | 50 mM Tris HCl, pH 8.0 |
| | 50 mM NaCl |
| | 2 mM DTT |
| Mono Q buffer B | 50 mM Tris HCl, pH 8.0 |
| | 1 M NaCl |
| | 2 mM DTT |
| Gel Filtration buffer | 50 mM Tris, pH 8.0 |
| | 50 mM NaCl |
| | 2 mM DTT |

**Expression and purification of seleno-methionine human *Ste20* OSR1 kinase**

Basically, the expression and purification procedure for seleno-methionine substituted human Ste20p OSR1 is carried out in the same fashion as wild-type. The important modification for this Se-Met procedure is to make sure that all the methioine must be substituted by seleno-methionine. To achieve such goal, methionine-free minimal media must be used substituted by seleno-methionine, together with methionine auxotrophic strain B834 (DE3) cells (Novagen). Even though the general procedure was similar, the growing rate for this auxotroph strain was much slower. This methionine auxotrophic strain B834 (DE3) cells transformed with the subcloned pHis-parallel vector was grown in minimal media with other 19 amino acids. A detailed recipe for preparing minimal media was described in **Table 2-2**. Unfortunately, however, Se-Met protein was not expressed enough for further purification, probably due to the fact that this human *Ste20p* OSR1 kinase construct must be expressed in the presence of ethanol-substituted medium for facilitating proper folding of target protein. The more detailed explanation for protein expression condition using ethanol was given in later part of this chapter (Result and Discussion). As a result, molecular replacement method was utilized to solve the crystal structure of OSR1 kinase, instead of using experimental phasing.

**Table 2-2 Recipe for minimal media to express Seleno-methionine protein**

| | |
|---|---|
| Ammonium sulfate | 1g/L |
| Adenine, guanosine, thymine and uracil | 125 mg/L of each nucleotide |
| Glucose | 10 g/L |
| 19 amino acids (except methioine) | 50 mg/L of each amino acid |
| Magnesium sulfate | 1 mM |
| Thiamine and biotin | 4 mg/L of each |
| Potassium phosphate monobasic | 4.5 g/L |
| Pottasium phosphate dibasic | 10.5 g/L |
| Sodiumchloride and sodium sulfate | 0.5 g/L |

**Crystallization of human *Ste20* OSR1 Kinase**

Every soluble construct produced was concentrated and screened for crystallization by utilizing commercially available sparse matrix kits (Emerald Biosciences and Hampton Research). The hanging-drop vapor diffusion method was used to grow crystals at 16 °C. Initial crystallization condition was identified by Hampton Screen kit I and extended for improving its quality and size by using home-made solution. One to five μl of protein was mixed with equal or varying amount of well solution (20% PEG 3350, 20mM Tri-Lithium citrate tetrahydrate and pH 8.0). The crystal appeared usually 2~3 days after set-up and grew up to dimension 50 X 60 X 300 in five days.

**Results and Discussion**

As mentioned above earlier, the successful expression and purification of target protein in crystallization trials usually presents a set of challenges that should be addressed from the various viewpoints. There are basically several points I focused upon throughout my experimental procedure.

First, expression system should be chosen for optimized target protein expression and solubility. In most cases, either *E.coli* or insect cell is utilized as a basic tool. It is well known that eukaryotic expression system like insect cell or yeast frequently turns out to produce more soluble and stable eukaryotic target protein, compared to prokaryotic system like *E.coli*. It is well known that efficient expression of eukaryotic target protein is frequently hampered by the low abundance of certain *E.coli* tRNAs that are abundant in eukaryotes. When the vector carrying the heterologous gene

is transformed into *E.coli* and target protein expression is iduced, the limited pool of low abundance tRNA is depleted and gives rise to translational arrest. Recently, a set of new *E.coli* strains including CodonPlus (Stratagene) and Rosetta (Novagen) was designed to address this codon-usage incompatibility issue. These new commercially available strains were equipped with pre-transformed plasmids encoding those low abundance tRNAs.

Secondly, the optimized construct should be generated in order to make sure that the expressed target protein is stable and suitable for crystallization trials. In order to find out which construct gives rise to the optimized protein behavior, a set of constructs should be tested for expression trials, considering other factors including expression system, expression conditions and various purification schemes.

Third, each construct tested for expression and purification trials might need additional fine-tuned optimization procedures towards successful protein crystallization. Even when soluble and stable target protein is available, it is often found that target protein is not crystallized. In this case, mild or radical changes should be considered in terms of construct design, expression system or purification procedures. For example, some mutant form of protein might be considered to produce more homogeneous sample or expression scheme such as growing condition may be modified to make sure that the expressed target protein is successfully crystallized to diffracting level.

**Expression of human *Ste20* OSR1 Kinase**

Prior to expression trials of human *Ste20* OSR1 kinase, basic sequence analysis was performed based upon the previous study (Chen, Yazicioglu et al. 2004). As shown

in **Figure 2-2**, this 527 amino acid-length *Ste20p* OSR1 kinase mainly consists of N-terminal kinase domain and C-terminal regulatory domain, which characterizes the Germinal Center Kinase subfamily (GCK).  The N-terminal extreme end of 17 amino-acid length is followed by regular kinase domain.  Right after the regular kinase domain, about 50 residues termed as PF1 domain is located and characterized as part of core kinase domain (Chen, Yazicioglu et al. 2004).  And about 100 amino-acid length of PF2 domain is located in the C-terminus of OSR1 kinase.  A recent study suggests that this PF2 domain functions as binding or oligomerization motif, without affecting the kinase activity itself (Chen, Yazicioglu et al. 2004).
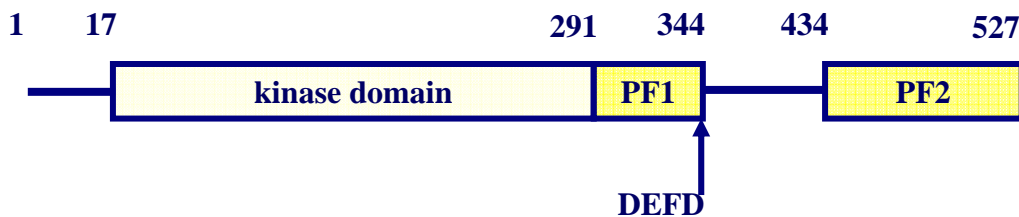


**Figure 2-2.  Domain Organization of human *Ste20* OSR1 kinase**

Even though it is not still clear how this protein kinase is activated, previous research suggests that two identified phsophorylation sites as shown in **Figure 2-3** (Thr 185 in activation loop and Ser 325 in PF1 domain) are involved in the activation mechanism by unknown mechanism (Anselmo, Earnest et al. 2006; Vitari, Thastrup et al. 2006).

## ⬛ Putative Activation mechanism of OSR1 kinase
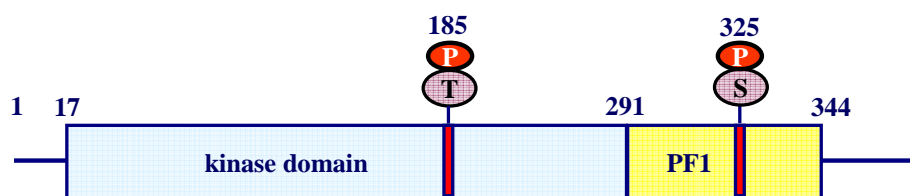


**Figure 2-3. Two major phosphorylation sites of human *Ste20* OSR1 kinase**

Based upon the current information on OSR1 kinase, a series of constructs was subcloned into pHis-parallel vector with N-terminal 6X histidine tag. As indicated in **Figure 2-4**, the generated constructs for prokaryotic or eukaryotic expression system include various fragments of human *Ste20p* OSR1 kinase. Every construct I designed (as shown in Figure 2-3) was tested for protein expression in *E.coli*. Initially, I used BL21 (DE3) for protein expression trials but the yield was not so high as other strain like Rosetta 2 (DE3) (Novagen). Since then, throughout the expression trials with *E.coli*, Rosetta 2 (DE3) was utilized for the reason I briefly discussed earlier. For the eukaryotic

expression trials, SF9 (Invitrogen) cells were used. For this SF9 cell trial, I tried two constructs (kinase domain with PF1 domain and kinase domain only).

Generally, all the constructs I designed were expressed from *E.coli*. Among those constructs, the expression level for the constructs containing 17-residue N-terminal insertion and the kinase domain core with or without PF1 domain was higher than the other constructs I designed. Interestingly, the expression level was much lower for the construct without the N-terminal 17 residues preceding kinase domain core. This N-terminal 17 amino acid sequence is thought to play an unknown but critical role in stabilizing this protein kinase folding. For the eukaryotic expression trials using SF9 cells, only kinase domain with PF1 domain was expressed well. The expression of kinase domain without PF1 domain in SF9 cells was almost absent. It is still unclear why kinase domain without PF1 domain was not expressed in poor yield from SF9 cells.

Following the extensive expression trials of every construct I designed for *E.coli* or insect cells, it was shown that the kinase domain with or without PF1 domain expressed from both systems is the most promising target for protein crystallization among other constructs, with regard to its solubility and stability. It thus follows that I focused on two constructs including the kinase domain with or without PF1 domain for further purification optimization by utilizing Rosetta 2 (DE3) and SF9 cell as expression system.

After substantial amount of purification optimization trials, however, it turned out that the final purified product of OSR1 kinase domain with PF1 domain from *E.coli* and insect cells has additional impurity. This construct containing kinase and PF1 domain expressed from *E.coli* and insect cells failed to be crystallized even after

extensive trials, probably owing to the observed sample heterogeneity.  For the reason mentioned just above, I was able to narrow my focus on only one construct expressed from *E.coli* finally.  The kinase domain without PF1 domain (except a few residues) expressed from Rosetta 2 (DE3) was purified without noticeable impurity.

Expression trials of human OSR1 in *E.coli &* SF9
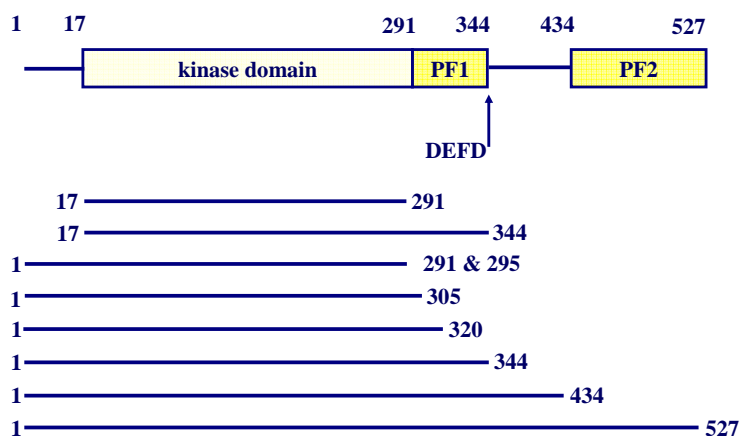


**Figure 2-4.  Design of constructs for expression trials in *E.coli* & insect cell**

Interestingly, however, it was found later that the OSR1 kinase construct without PF1 domain expressed from *E.coli* strains continued to be precipitated at low level during the purification steps, especially after Ni affinity column.  After I failed to crystallize this construct (kinase domain without PF1 domain), I tried extensive optimization schemes

including different growing conditions or other variations to improve the stability of this target protein construct.

In other words, despite the fact that this target protein was soluble and stable enough for expression, purification and even concentration up to 10 mg/ml range, some population of this target protein continued to be unstable and vulnerable to aggregation during the whole purification steps. In fact, however, it is no wonder that the heterologous protein expression frequently gives rise to misfolding coupled to stability problem, probably due to the difference in the folding machinery provided by prokaryotic and eukaryotic system.

However, it is well known that when *E.coli* is exposed to a variety of stresses such as temperature upshift, contact with organic solvents and the accumulated misfolded proteins, the synthesis of 20 ~ 30 heat shock proteins is transiently upregulated to repair cellular damage (Neidhardt 1987; VanBogelen, Kelley et al. 1987; Thomas and Baneyx 1997). Although the cellular function of many hsps remains unclear, several are known to be ATP-dependent proteases and molecular chaperones, which help nascent and partially folded proteins reach a proper final confirmation or specific location (Thomas and Baneyx 1997).

Recent studies show that the proper folding of aggregation-prone recombinant proteins in *E.coli* may be facilitated by co-overexpressing specific molecular chaperones or by growing the cells in the presence of ethanol or other agents that can boost the production of all heat-shock proteins (hsps) (Thomas and Baneyx 1997). By exploiting the fact that proper folding of heterologous protein expression may be facilitated via induction of molecular chaperones by culturing the cells in the presence of ethanol, I

designed a set of expression trials for improving solubility of target protein.  About 1% to 3% range of ethanol was introduced to LB culture medium, not to small volume of starter culture, as suggested by previous study (Thomas and Baneyx 1997).  3% ethanol-substituted LB culture medium gave the best result, producing more stable protein sample.

Since the introduction of ethanol-substituted growing medium into expression procedure, any type of aggregated target protein was never formed and was also expressed in high level.  Finally, a subtle problem of protein stability related to partial misfolding was resolved with the help of molecular chaperones induced by ethanol supplementation.

**Purification of human *Ste20* OSR1 Kinase**

As explained above in [Methods and Procedures], standard purification steps were employed to obtain the final homogeneous sample.  The entire protein purification steps were finished usually within two days.  The result of each purification step was shown in a series of pictures from Figure 2-5 to Figure 2-11.

After thawing the frozen cells, the lysozyme and bezonase nucleases were utilized to help break the cells.  Following the sonication step, the cell extract was applied to ultracentrifugation for removing cell debris.  Interestingly, but not surprisingly, the protein expression derived from ethanol-added culture media produced clearer and more transparent supernatant solution than ethanol-free expression.

For the Ni affinity chromatography, step gradient was used to elute this target protein by applying 100% 1M imidazole-containing buffer B.  The eluted sample was

diluted immediately with MonoQ buffer A (from anion-exchange chromatography) to prevent precipitation due to high salt concentration. This diluted sample was then digested with TEV protease overnight during dialysis step to remove N-terminal 6X histidine tag. During this TEV protease cleavage step, the OSR1 kinase domain expressed from ethanol-free culture medium was usually precipitated even at low level. However, ethanol-containing culture medium produced precipitation-free target protein over all the purification steps. Also, it should be mentioned that there was no difference in protein behavior between 1~295 construct (kinase domain without PF1 domain) and 1~344 construct (kinase domain with PF1 domain).

The TEV protease cleavage step was followed by anion-exchange chromatography (MonoQ: Amersham) for further purification. After loading protein sample, gradual gradient was applied by using 1M high salt (NaCl) – containing buffer B.

In the case of kinase domain without PF1 domain (1~295), the major population of the target protein did not bind to MonoQ column as shown in **Figure 2-5**. Most of the target protein (1~295) was eluted from flow-through section prior to gradient and pooled for the next purification step. In fact, some of target protein (1~295) was also eluted from early part of gradient and tested for the next step. However, target protein eluted from the early gradient of anion-exchange chromatography was not stable, although it was highly pure based upon SDS-PAGE analysis. The eluted samples from gradient were collected and processed for TEV protease cleavage, but they were precipitated completely in a short period of time.

In the case of kinase domain with PF1 domain (1~344), most of the target protein did bind to MonoQ column as shown in **Figure 2-8**. The target protein (1~344) was

eluted from the early part of gradient. It was also highly homogeneous sample based upon SDS-PAGE analysis.
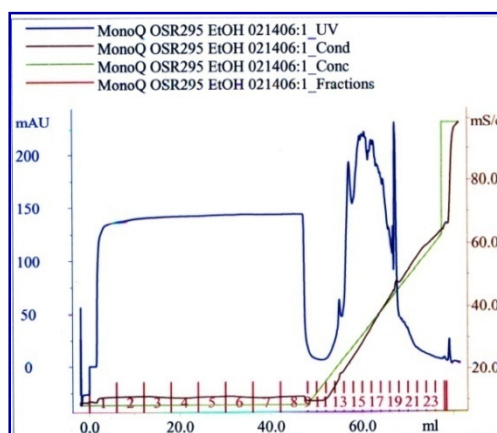


**Figure 2-5.** **Purification of OSR1 kinase (1~295) by anion- exchange chromatography**
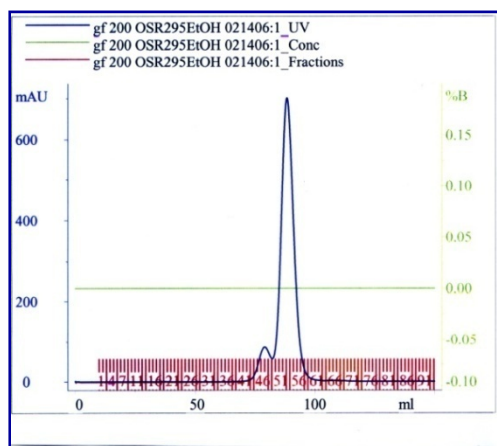


**Figure 2-6. Purification of OSR1 kinase (1~295) by size-exclusion chromatography**
The small peak shown right before the major peak was not collected for the next step of concentration.
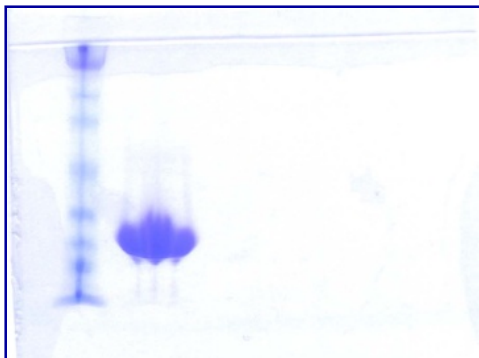
**Figure 2-7.  SDS-PAGE gel picture for the purified OSR1 kinase (1~295)**
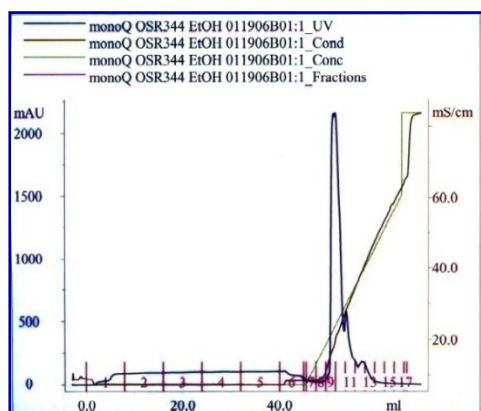


**Figure 2-8.    Purification of OSR1 kinase (1~344) by anion-exchange chromatography**
The eluted samples from gradient were collected and processed for TEV protease cleavage.  There was only low level of protein collected from the flow-through part.
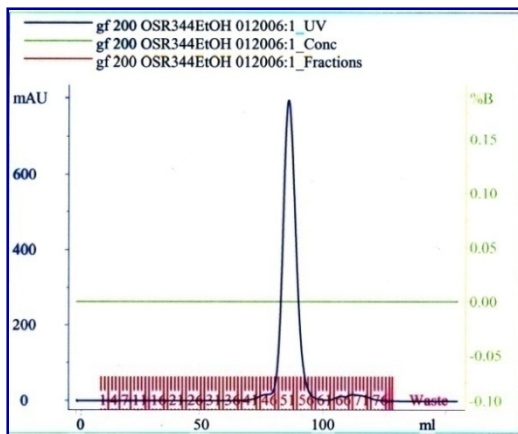
**Figure 2-9. Purification of OSR1 kinase (1~344) by size exclusion chromatography**
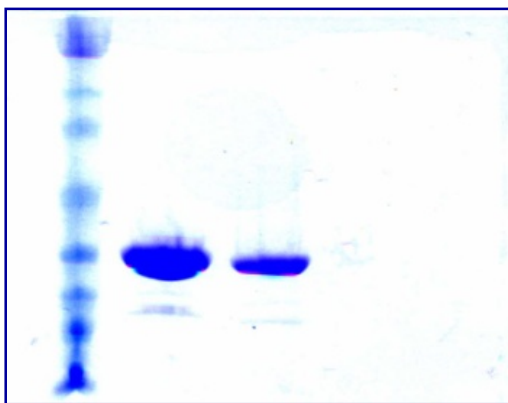


**Figure 2-10. SDS-PAGE gel picture for the purified OSR1 kinase (1~344)**
    Two loaded samples are duplicated for experiment. The impure band was shown clearly right below the purified target protein.

Anion-exchange chromatography was followed by size-exclusion chromatography for both constructs (1~295 and 1~344). Both samples produced one single peak from gel filtration, indicating that two proteins formed monomer in solution as shown in **Figure 2-6 and Figure 2-9**. Following anion-exchange chromatography, target proteins were collected and concentrated up to 10~12 mg/ml. Even during this concentration step, protein samples of both constructs (1~295 and 1~344) derived from ethanol-free culture medium generated some precipitation. Protein samples from ethanol-added medium, however, showed no sign of instability. As shown in **Figure 2-7**, 1~295 construct turned out to be homogeneous purified sample based upon SDS-PAGE analysis. In contrast, as indicated in **Figure 2-10**, one impure band from 1~344 construct was never removed, even after trying other modified purification methods extensively. Based upon concentration trials, the solubility of both constructs seemed to be similar. Finally, both protein samples were concentrated and tested for crystallization trials. The final protein sample buffer contains Tris 50mM pH=8.0, NaCl 50mM.

**Crystallization of human *Ste20* OSR1 Kinase**

As briefly mentioned above in [Methods and Procedures], initial extensive crystallization screening for both constructs (1~295 and 1~344) was made by using commercially available kits manufactured from Hampton or Emerald. For both constructs, various additives such as detergents or nucleotide-analogs were utilized for crystallization trials. 1~344 construct didn't provide reproducible crystallization even after extensive screening, although there was an initial condition (not reproducible) as shown in Figure **2-14**. Only 1~295 construct was crystallized in the presence of 0.1% β-

octylglucoside, 2mM AMP-PNP, 5mM MgCl, as shown in **Figure 2-11**. The initial crystallization condition for 1~295 OSR1 construct contains 50mM Potassium dihydrogen phosphate, 20% w/v PEG8000. The initial crystal took the form of thin rod about 50 μm long. Since then, an extensive amount of efforts was focused upon expansion and improvement of initial crystallization condition.

After the extensive efforts, the improved crystals were grown as shown in **Figure 2-12 and Figure 2-13**. This improved crystallization condition contains 200mM tri-Lithium citrate, 20mM Taurine, 20% w/v PEG 3350, also in the presence of 0.1% β-octylglucoside, 2mM AMP-PNP, 5mM MgCl. These crystals appeared usually 3 days after its set-up. And these crystals were grown for several days at 20°C. Also, in order to increase the size of crystals, the bigger drops for protein and well solution (2~4μl + 2~4μl) were used and it gave me the resulting bigger protein crystal as shown in **Figure 2-13**, which was utilized for synchrotron data collection to solve the crystal structure.
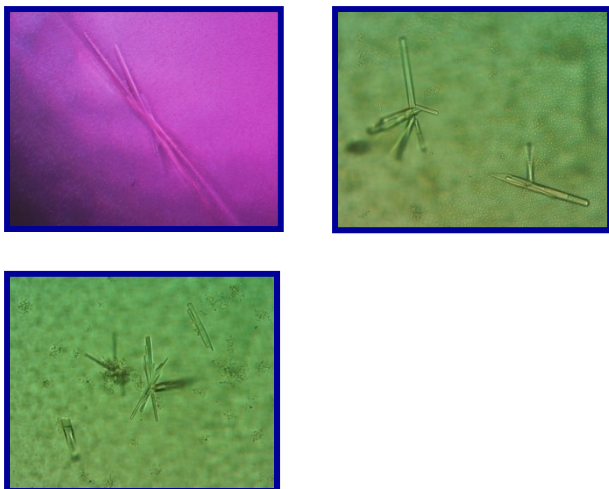
**Figure 2-11.  Initial crystals of human *Ste20* OSR1 kinase (1~295)**
The crystals were grown by hanging drop vapor diffusion method.  This initial condition contains 50mM Potassium dihydrogen phosphate, 20% w/v PEG8000.  These crystals were grown for 3 days at 20°C.  The length of these crystals is about 50 μm.
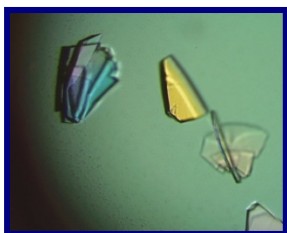


**Figure 2-12.  Improved crystals of human *Ste20* OSR1 kinase (1~295)**
The crystals were grown by hanging drop vapor diffusion method.  This improved crystallization condition contains 200mM tri-Lithium citrate, 20mM Taurine, 20% w/v PEG 3350.  These crystals were grown for 4 days at 20°C.
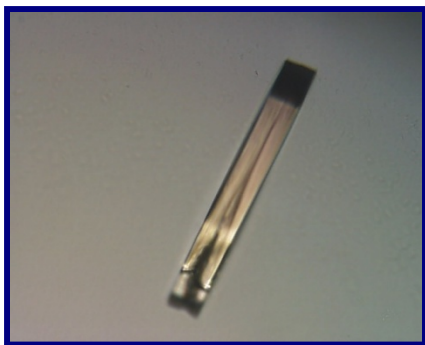
**Figure 2-13.  Improved crystals of human *Ste20* OSR1 kinase (1~295) used for data collection**
The crystals were grown by hanging drop vapor diffusion method.  200mM tri-Lithium citrate, 20mM Taurine , 20% w/v PEG 3350.  These crystals were grown for 4 days at 20°C.  Compared to the crystals shown in Figure 2-12, these crystals were grown from a series of bigger drops (2~4 µl of protein solution plus 2~4 µl of well solution).



**Figure 2-14.  Initial crystallization of human *Ste20* OSR1 kinase (1~344)**
The crystals were grown by hanging drop vapor diffusion method.
This crystal condition contains 30% PEG4000, 0.1M TrisHCl pH 8.5, 0.2M Sodium Acetate.

**Expression and purification of seleno-methionine human *Ste20* OSR1 kinase**

As described in the earlier section [Methods and Procedures], seleno-methionine substituted protein was tried for expression from methionin auxotroph B834 (DE3) cells. Unfortunately, however, the seleno-methionine substituted protein was expressed in a very low level, which makes further purification step and crystallization virtually impossible. It thus follows that I decided to go for heavy atom screening in order to figure out phase problem. But the heavy atom screening also failed. Since then, as the last method I can rely upon to solve a phase problem, molecular replacement began to be considered, as explained in great detail in the next chapter.

# Chapter 3: Structure determination of human *Ste20* OSR1 kinase domain

## Introduction

As discussed in Chapter 2, the crystals of the human *Ste20p* OSR1 kinase domain were improved and optimized to a size of about 50μm X 60μm X 300μm. However, the crystals did not diffract well to high resolution range (about 2.8 Å) using in-house rotating anode source. The crystal structure of human *Ste20p* OSR1 kinase domain was determined from X-ray diffraction data set, utilizing the high brightness X-ray beams generated by the Advanced Photon Source (APS) at Argonne, Illinois. The data set of OSR1 kinase domain was collected over 150º with a rotation range of 0.25º for each 8-second exposure.

The crystal structure of the human *Ste20p* OSR1 kinase domain was solved by molecular replacement *Phaser* (Read 2001; Storoni, McCoy et al. 2004; McCoy, Grosse-Kunstleve et al. 2005). The checkpoint kinase Chk1 (PDB code 1NVR) (Zhao, Bower et al. 2002) and *Ste20p* TAO2 kinase (PDB code 1U5R) (Zhou, Raman et al. 2004) were utilized as search models for N-terminal lobe and C-terminal lobe of OSR1 kinase domain, respectively, after these models were modified by CHAINSAW to mutagenize bulky side chain such as Trp or Tyr into Ser or Ala and some regions of those models were deleted to avoid packing problem associated with molecular replacement. The utilization of both subdomain fragments as search models for molecular replacement

yielded clear rotation and translation peaks. After extensive manual modeling into calculated electron density map for the deleted and modified residues, the improved model was again fed onto program *Phaser* for additional cycle of molecular replacement. The structur of OSR1 kinase domain ws refined with iterative cycles of manual modeling and refinements by utilizing *REFMAC5*.

## Materials and Procedures

### Data Collection and Data Reduction

Data collection for human *Ste20p* OSR1 kinase domain was performed on a charge-coupled device (CCD) detector at 90K in the APS (19BM beam line), with the help of Dr. Diana Tomchick.

The HKL package (Denzo and Scalepack) was utilized for data processing including autoindexing, integration of diffraction intensities, data correction, scaling and post refinement. To identify the correnc space group of OSR1 kinase data set, systematic absences were analyzed using the SCALEPAK module within HKL package. Due to the clear patters of absences in reflections, the space group of this data set was decided ($P2_12_12_1$) as shown in **Table 3-1**. In addition, I tried all the possible orthorhombic spacegroups ($P222_1$, $P2_12_12$, $P2_12_12_1$, $P222$) were tested using *Phaser* to find productive rotation and translation solutions.

In order to estimate the number of molecules per asymmetric unit, Matthew's coefficient ($V_M$) was utilized. $V_M$ is calculated as the ratio between the unit cell volume and the molecular weight of protein contained in the unit cell.

$$V_M = V / nM$$

where V is the volume of the crystal's unit cell, $n$ is the number of copies of protein monomer in the unit cell and M is the molecular weight of the protein. For most of protein crystals, $V_M$ values lie between 1.7 and 3.5 $\text{Å}^3 \, \text{Da}^{-1}$. For the OSR1 kinase crystal with space group $P2_12_12_1$, 4 molecules per aymmetric unit corresponds to 16 molecules per unit cell. Given that the molecular weight of OSR1 kinase domain (M) is 33170 Da and the volume of the crystal's unit cell (V) is 1263076 $\text{Å}^3$, the calculated $V_M$ would be 2.38, which lies within the reasonable distribution limits.

**Table 3-1 Intensities of systematic absences (*h00, 0k0, 00l*:*hkl* ≠ *2n*)**

| h | k | l | Intensity | Sigma | I/Sigma |
|---|---|---|-----------|-------|---------|
| 0 | 0 | 11 | 11.2 | 1.0 | 11.4 |
| 0 | 0 | 13 | -2.5 | 1.0 | -2.6 |
| 0 | 0 | 15 | 23.9 | 0.9 | 26.6 |
| 0 | 0 | 17 | 3.2 | 0.9 | 3.7 |
| 0 | 0 | 19 | -2.4 | 0.7 | -3.3 |
| 0 | 3 | 0 | 11.6 | 2.5 | 4.5 |
| 0 | 5 | 0 | 2.1 | 2.0 | 1.1 |
| 0 | 7 | 0 | -8.2 | 3.0 | -2.7 |
| 0 | 9 | 0 | -10.4 | 4.3 | -2.4 |
| 0 | 11 | 0 | -1.8 | 4.9 | -0.4 |
| 0 | 13 | 0 | -2.9 | 6.0 | -0.5 |
| 0 | 15 | 0 | -27.1 | 7.3 | -3.7 |
| 0 | 19 | 0 | -49.5 | 10.0 | -4.9 |

| 0 | 21 | 0 | −52.8 | 17.6 | −3.0 |
|---|---|---|---|---|---|
| 0 | 23 | 0 | −75.3 | 13.9 | −5.4 |
| 0 | 25 | 0 | −14.4 | 15.9 | −0.9 |
| 0 | 27 | 0 | 517.1 | 76.8 | 6.7 |
| 0 | 29 | 0 | 145.7 | 20.8 | 7.0 |
| 0 | 31 | 0 | −41.4 | 18.8 | −2.2 |
| 0 | 33 | 0 | −10.8 | 18.4 | −0.6 |
| 0 | 35 | 0 | −11.7 | 19.2 | −0.6 |
| 0 | 37 | 0 | −36.4 | 20.5 | −1.8 |
| 0 | 39 | 0 | −0.9 | 18.3 | 0.0 |
| 0 | 41 | 0 | 5.0 | 19.8 | 0.3 |
| 0 | 43 | 0 | −18.3 | 19.0 | −1.0 |
| 0 | 45 | 0 | −16.8 | 20.0 | −0.8 |
| 3 | 0 | 0 | 8.3 | 2.2 | 3.8 |
| 5 | 0 | 0 | −0.3 | 3.7 | −0.1 |
| 7 | 0 | 0 | 2.2 | 5.1 | 0.4 |
| 9 | 0 | 0 | 3.8 | 7.0 | 0.5 |
| 11 | 0 | 0 | 12.3 | 11.1 | 1.1 |

| 13 | 0 | 0 | −9.7 | 10.6 | −0.9 |
|----|---|---|------|------|------|
| 15 | 0 | 0 | 55.0 | 12.9 | 4.3 |
| 17 | 0 | 0 | 1.1 | 15.3 | 0.1 |
| 19 | 0 | 0 | 1.5 | 18.6 | 0.1 |
| 21 | 0 | 0 | −35.5 | 20.7 | −1.7 |
| 23 | 0 | 0 | −15.3 | 20.7 | −0.7 |
| 25 | 0 | 0 | 32.4 | 22.5 | 1.4 |
| 27 | 0 | 0 | −26.7 | 24.6 | −1.1 |
| 29 | 0 | 0 | −9.7 | 26.9 | −0.4 |
| 31 | 0 | 0 | −13.9 | 28.9 | −0.5 |

**Introduction to Macromolecular Crystallography: The Phase Problem**

In the diffraction experiment, the intensities of waves scattered from planes (denoted by *hkl*) in the crystal are measured. The basic aim of crystallographic method is to record the diffraction pattern from a crystal using X-ray source and to utilize those diffraction intensities to understand molecular structure in the crystal. The element of the diffraction spectra (the structure factors) from a crystal may be written as follows (Taylor 2003).

$$\boldsymbol{F_h} = V \, \Sigma \, \boldsymbol{f_j} \exp i2\pi(\boldsymbol{r_j} \cdot \boldsymbol{h})$$

Where $\boldsymbol{F_h} = \boldsymbol{F_{hkl}}$ = structure factor, $V$ is the volume of unit cell, $f_j$ is the scattering factor of all the atoms (*j*) in a unit cell, $r_j$ is the coordinates $x_j$, $y_j$, $z_j$ of the $j^{th}$ atom, $\boldsymbol{h}$ is the reciprocal lattice vector. This equation tells us that given the atomic coordinates of a crystal, we can calculate the intensities at each reciprocal lattice point in the diffraction pattern of the crystal.

The amplitude of the wave $|F_{hkl}|$ is proportional to the square root of the intensity measured on the detector. In order to calculate the electron density at a position (*xyz*) in the unit cell of a crystal, we perform the following summation over all the hkl planes. It means that electron density at (*xyz*) is equal to the sum of contributions to the point (*xyz*) of waves scattered from plane (*hkl*) whose amplitude depends upon the number of electrons in the plane, added with the correct relative phase relationship as follows (Taylor 2003),

$$\boldsymbol{\rho(xyz)} = 1/V \, \Sigma \, |\boldsymbol{F_{hkl}}| \exp(i\alpha_{hkl}) \exp(-2\pi ihx + ky + lz),$$

where $V$ is the volume of the unit cell and $\alpha_{hkl}$ is the phase associated with the structure–

factor amplitude $|F_{hkl}|$. According to this equation, we can calculate directly the value of

the electron density, for the given structure factors for any point x, y, z in the unit cell.

However, $F_{hkl}$ is a vector quantity, which is complex number consisting of phase and

amplitude. The measurement of intensities can provide the amplitudes of the structure

factors, but the phase information is lost. Thus, a variety of methods (molecular

replacement, isomorphous replacement, anomalous replacement) to overcome this phase

problem has been developed and utilized (Taylor 2003).


**Solving the phase problem by Molecular Replacement (MR)**

When a homology model is available, crystallographer can utilize the phases

from structure factors of a known protein structure as initial estimates of phases for a new

protein. If this method is feasible, the structure of a new molecule may be solved from a

single native data set. The known protein in this case is termed as a phasing model, and

the method, which entails calculating initial phases by placing a model of the known

protein in the unit cell of the new protein, is called molecular replacement (MR).

The basic goal of MR is to orient and position the search model, such that it

coincides with the position of the unknown protein in the crystal. A total of six

parameters need tobe determined – 3 anlges and 3 translational elements.

For example, the mammalian serine proteases, trypsin, chymotrypsin, and

elastase, are very similar in structure and conformation. If a new mammalian serine

protease is discovered, and sequence homology with known proteases suggest that this

new protease is similar in structure to known ones, one of the known proteases might be used as a phasing model for determining the structure of the new protein.

As a rule of thumb, a sequence identity >25% is normally required and r.m.s.deviation of <2.0 Å between αC atoms of the model and the final new structure, even though there are some exceptions. Patterson methods are usually used to obtain first the orientation of the model in the new unit cell and then the translation of the correctly oriented model relative to the origin of the new unit cell as shown in **Figure 3-1**.
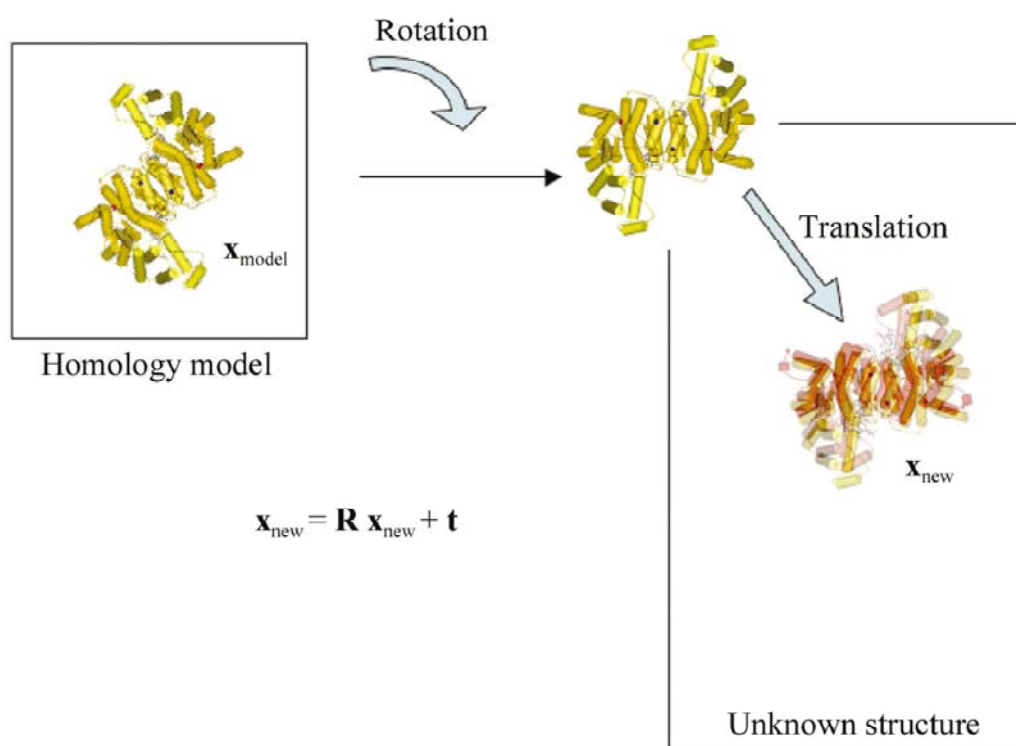


**Figure 3-1. The process of molecular replacement.**
This figure was adapted from (Taylor 2003).

**Molecular Replacement with the Protein Kinase**

In the case of solving a new protein kinase structure with molecular replacement, it presents a unique challenge in terms of its conformational variance. It is well known that the core kinase domain structure and its sequence are highly conserved. In many cases, however, the molecular replacement with protein kinase (for a new structure) turned out to be unsuccessful. We can suggest some rationales for difficulty in molecular replacement with protein kinase. The most basic feature of core protein kinase is that it consists of two lobal subdomains. Even though the protein kinase is called "domain", there are two separate subdomains in terms of its folding and functionality as shown in **Figure 3-2.** Each subdomain is classified into two lobes, N-terminal lobe and C-terminal lobe. Whereas the N-terminal lobe is basically involved in nucleotide binding, the C-terminal lobe is the main region for substrate binding. These two lobes connect to each other in different manners, by a variety of relative lobal rotation. Also, depending on the "active" or "inactive" states of kinase activity, the entire kinase domain reveals a wide range of lobal interactions, which prepares kinse domain ready for nucleotide or substrate binding. In addition, whereas C-terminal lobe is thought to be relatively "static", N-terminal lobe shows more variant folds and movements. As a result, structure alignment from C-terminal lobe reveals highly conserved homology whereas N-termial lobal structure alignment usually turns out to be "variant" and "perturbed". In other words, even high sequence homology among the given protein kinase family or group is not coupled to high structure homology. Those factors just mentioned earlier allow us to discard the long-held notion that protein kinase core domain is highly conserved in terms of sequence and structure. Furthermore, this realization of extreme diversity of protein

kinase core helps us understand why in many cases molecular replacement with protein kinase for new structure is failed. It is no wonder that molecular replacement with protein kinase frequently cannot find any solution, especially when the entire kinase is used as search model.
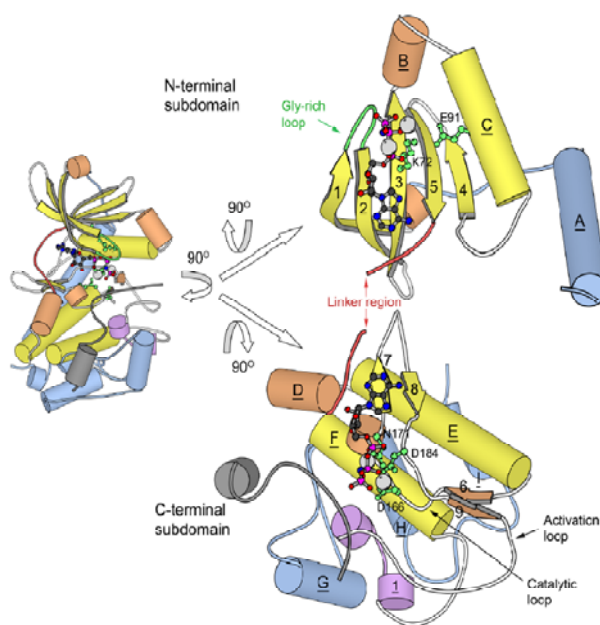


**Figure 3-2. Two Views of the Structure of PKA**
The structure of a kinase domain consists of two subdomains: a small, mainly β-sheet N-terminal subdomain, and a larger, primarily helical C-terminal subdomain. ATP and metal ions are bound in the cleft between the two lobes. Substrate is bound mainly via the C-terminal lobe. The left-side small kinase domain depicts PKA as a standard model for the protein kinase. N-terminal subdomain is usually characterized by 5 β-sheets and αC-helix whereas C-terminal lobe contains a well-conserved catalytic loop beginning with RD followed by helices E and F. This model diagram was adapted from (Scheeff and Bourne 2005).

**Molecular Replacement for OSR1 kinase domain**

As discussed in earlier chapters, SeMet-substituted OSR1 kinase domain was not produced from *E-coli*. Also, the extensive trials for heavy atom screening were not successful either. Although molecular replacement remained a feasible tool for solving the crystal structure of OSR1 kinase domain, this method presented its unique challenges with regard to protein kinase as mentioned above. In order to address the molecular replacement problems associated with protein kinases, different approach was utilized, in which the whole kinase domain was considered as two separate lobes. Instead of using the entire kinase domain as search model, two separate search models for each lobe was selected to solve this structure. In other words, one protein kinase domain is now regarded as complex structure consisting of two proteins, not single protein. In this case of dealing with OSR1 kinase for molecular replacement, this approach means that there exist smaller 8 molecules of two kinds, since there are 4 molecules of one kind in the asymmetric unit, according to Matthew Coefficient analysis, as desribed in **Figure 3-3**.

By utilizing SCOP (<u>S</u>tructural <u>C</u>lassification <u>O</u>f <u>P</u>roteins) database and BLAST (<u>B</u>asic <u>L</u>ocal <u>A</u>lignment <u>S</u>earch <u>T</u>ool), search models for N-terminal and C-terminal lobes were searched separately and tested for molecular replacement towards OSR1 kinase domain structure with program *Phaser*. Since it is thought that larger C-terminal lobe must contribute more to the entire phase calculation than smaller N-terminal lobe, search model for C-terminal was chosen first and tried for initial 4 molecules. After it was found that 4 molecules for C-terminal lobe was phased with a chosen model based upon the analysis of *Phaser*, additional 4 molecules for N-terminal lobe with another model

was then tried for complete phase. With about more than 40 N-terminal and C-terminal-lobe search models, extensive combinations of trials were tested for molecular replacement.

However, such trials with intact models didn't work initially. At this stage, I realized that there might be some additional modifications I can try on the search probes, after it was thought that conformational diversity in the protein kinase wasn't still addressed in a proper manner.

The first modification is simply based upon the fact that the backbone of search model, not the side chain, should be focused to reduce model bias associated with molecular replacement. I decided to utilize program *Chainsaw* to modify (mutagenize) bulky amino acid side chains in search model into polyalaine (much less bulky) residues, focusing on the main chain model, based upon the pairwise sequence alignment between a given search model and OSR1 kinase.

The second modification I introduced onto the search model is originated from the basic notion that N-terminal and C-terminal lobes in the protein kinase are actually intertwined to each other in terms of three-dimensional folding. It means that some portion of N-terminal lobe was extended further to overlap with part of C-terminal lobe with regard to Z-axis, probably due to some rotational rearrangement of N-terminal lobe. Thus, in addition to mutagenizing those bulky amino acid residues to focus on main chain, some of so-called "intertwined" residues were deleted from each N-terminal and C-terminal search probe, based upon the location of the targeted residues **Figure 3-4**. A wide range of residues were deleted from both N-terminal and C-terminal search models to generate an improved search probe for protein kinase molecular replacement,

depending upon the unique topology of each chosen model. Sometimes, various models were generated from one model, depending upon various regions of deleted residues.

Following the extensive trials employing a set of modified search probes as discussed above, the combination of the checkpoint kinase Chk1 (PDB code: 1NVR) as N-terminal lobal model and *Ste20* TAO2 (PDB code: 1U5R) kinase as C-terminal lobal model sucessfully found the complete solution for molecular replacement. In this case, about 10% of search model was deleted and used for molecular replacement. The sequence homology is about less than 30 % for N-terminal lobal region and approximately 40 % for C-terminal lobal model. After finding a phase solution, some model building was made onto the clear density over more than 30 residues. This improved model produced by the extensive modeling was utilized again for a better solution in molecular replacement **Figure 3-4**. After several rounds of manual modeling and molecular replacement, the density map was improved and $R_{free}$ value was reduced to 36.1 %.

# MR strategy for OSR1 KD (4 mols in asu)

Conventional MR approach
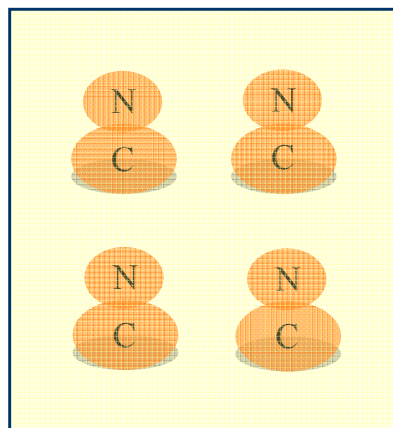
OSR MR approach



4 monomers
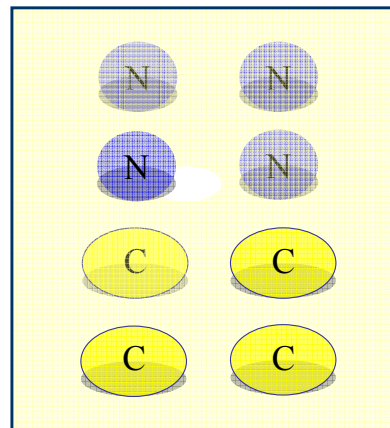
8 monomers

**Figure 3-3.  Molecular Replacement Strategy for OST1 kinase domain**
For molecular replacement, the kinase domain was divided into subdomains.  Single kinase domain was regarded as protein complex consisting of N-terminal lobal and C-terminal lobal proteins.  Instead of using single kinase domain as search probe, each model for N-terminal and C-terminal lobe was separately chosen from SCOP database. Due to the larger size associated with more contribution to entire phase, C-terminal lobal search probe was used first for finding solution and then N-terminal lobal model was searched.
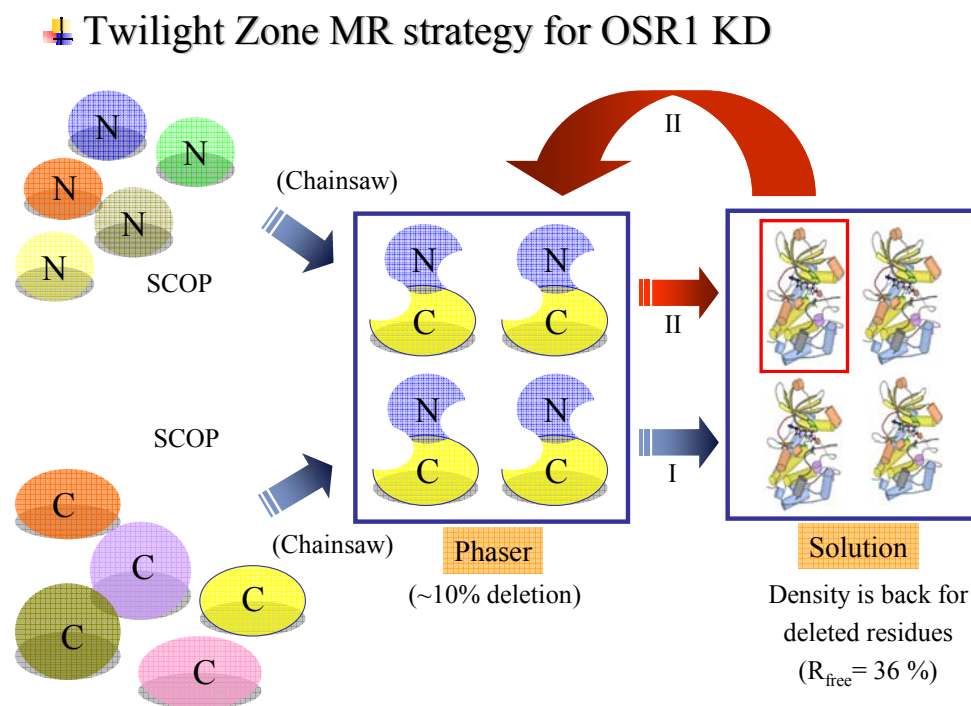
**Figure 3-4. The Basic Scheme for Twilight Zone Molecular Replacement in the structural determination of OSR1 kinase domain**

About 10% of residues were deleted from each chosen model to prevent lobal conflict derived from intertwined kinase folding. After solution was found, the density map allowed extensive modeling over 30 missing residues. The improved model was utilized again for additional run of molecular replacement.

**Model Building and Refinement**

As mentioned above, right after progam *Phaser* found the initial solution for OSR1 kinase domain by employing two different search probes for N-terminal and C-terminal lobes, the interpretable density map was produced to allow for manual modeling. Since some (about 10 %) residues from the initial model were deleted and bulky side chains such as tryptophan or tyrosine were also modified into serine or alanine, extensive manual modeling was required onto the calculated electron density map ($2F_o$-$F_c$). The improved OSR1 model after manual modeling was again fed onto molecular replacement using program *Phaser*. After several rounds of manual modeling and molecular replacement, the model was then applied to rigid-body and restrained refinement in *REFMAC5*. Due to thet fact the there are 4 molecules in the asymmetric unit, manual modeling involved in more than 1100 amino acids was challenging and exhaustive process. Following the initial refitting of most residues targeted for manual modeling, additional several cycles of refinement and modeling were performed.

In general, the electron density map from C-terminal portion was better than one from N-terminal lobe. Especially, E helix and F helix regions in C-terminal lobe gave most clear density map whereas some loop areas from both of two lobes produced poor-quality density map as expected. In order to improve model quality, various parameters such as contact and stereochemistry were checked using PROCHECK and MOLPROBITY. During the refinement steps, TLS MD (<u>T</u>ranslation/<u>L</u>ibration/<u>S</u>crew Motion <u>D</u>etermination) was also utilized and free *R* value was reduced to 26.7 % as indicated in **Table 3-2**.

**Table 3-2 Data collection and Refinement Statistics**

| | OSR1(kinase domain) |
|---|---|
| Space group | $P2_12_12_1$ |
| Unit cell dimensions (Å) | $a = 74.18$, |
| | $b = 104.42$, |
| | $c = 162.67$ |
| | $\alpha = \beta = \gamma = 90$ (°) |
| Wavelength (Å) | 0.97899 |
| Resolution range (Å) | $50.0 \sim 2.254$ |
| Unique reflections | 59432 |
| Multiplicity | 5.5(4.4) |
| Completeness (%) | 98.2(91.9) |
| $R_{merge}$ [a] (%) | 8.2(45.6) |
| Intensities I/σ(I) | 23.2(2.9) |
| $R_{work}/R_{free}$ [b] (%) | 19.5(26.7) |
| Non-H protein atoms | 9443 |
| Waters | 512 |
| No. of reflections in refinement ($F>2\sigma$) | 56330 |
| r.m.s.d. in bond length (Å) | 0.016 |

| | |
|---|---|
| r.m.s.d. in bond angles ( ° ) | 1.665 |
| Average $B$-values ($\text{Å}^2$) | 38.6 |

[a]$R_{\text{merge}} = \Sigma|(I_{\text{hkl}}) - <I>|/\Sigma(I_{\text{hkl}})$, where $I_{\text{hkl}}$ is the integrated intensity of a given reflection.

[b]$R = (\Sigma|F_{\text{o}} - F_{\text{c}}|)/( \Sigma F_{\text{o}})$, where $F_{\text{o}}$ and $F_{\text{c}}$ are observed and calculated structure factors, respectively.

# Chapter 4: Structure analysis of human *Ste20* OSR1 kinase domain

**Sequence Analysis**

*Ste20* OSR1 kinase belongs to GCK-VI (Germinal Center Kinase) subfamily, which is characterized by N-terminal kinase domain followed by C-terminal non-catalytic regulatory domain.

The sequence of OSR1 kinase domain reveals the typical, characteristic features, which are well conserved in the protein kinase family. Those features were discussed in great detail in the previsou chapter.

Sequence alignment analysis among GCK-VI subfamily members shows more than 90% homology in the protein kinase domain as indicated in **Table 4-1**. Due to high homology more than 90% in GCK-VI subfamily, it is difficult to find unique features about *Ste20* OSR1 kinase domain from this sequence analysis. As of analyzing this crystal structure of OSR1 kinase domain, no structure from GCK-IV subfamily has been reported.

As mentioned earlier in chapter 2, the entire Ste20 kinase group shares its unique signature sequence located in the C-terminal end of activation segment.

```
OSR1_Frog_    WSINRDDYELQEVIGSGATAVVQAAFCAPKKEKVAIKRINLEKCQ
OSR1_Human    WSINRDDYELQEVIGSGATAVVQAAYCAPKKEKVAIKRINLEKCQ
OSR1_Mouse_   WSINRDDYELQEVIGSGATAVVQAAYCAPKKERVAIKRINLEKCQ
SPAK_Human_   WPICRDAYELQEVIGSGATAVVQAALCKPRQERVAIKRINLEKCQ
SPAK_Mouse_   WPICRDAYELQEVIGSGATAVVQAALCKPRQERVAIKRINLEKCQ
SPAK_Rat_     WPICRDAYELQEVIGSGATAVVQAALCKPRQERVAIKRINLEKCQ
FRAY_Droso    WPNSKDDYELRDVIGVGATAVVHGAYCIPRNEKCAIKRINLEKWN

cons          *.   :* ***::*** ******:.* * *::*: ********* :


OSR1_Frog_    TSMDELLKEIQAMSQCHHPNIVSYYTSFVVKDELWLVMKLMSGGS
OSR1_Human_   TSMDELLKEIQAMSQCHHPNIVSYYTSFVVKDELWLVMKLLSGGS
OSR1_Mouse_   TSMDELLKEIQAMSQCHHPNIVSYYTSFVVKDELWLVMKLLSGGS
SPAK_Human_   TSMDELLKEIQAMSQCSHPNVVTYYTSFVVKDELWLVMKLLSGGS
SPAK_Mouse_   TSMDELLKEIQAMSQCSHPNVVTYYTSFVVKDELWLVMKLLSGGS
SPAK_Rat_     TSMDELLKEIQAMSQCSHPNVVTYYTSFVVKDELWLVMKLLSGGS
FRAY_Droso    TSMDELLKEIQAMSSCFHENVVTYHTSFVVREELWLVLRLLEGGS

cons          **************.* * *:*:*:*****::*****::*:.***


OSR1_Frog     VLDIIKHIIARGEHKNGVLDEPSIATMLKEVLEGLEYLHKNGQIH
OSR1_Human    VLDIIKHIVAKGEHKSGVLDESTIATILREVLEGLEYLHKNGQIH
OSR1_Mouse    VLDIIKHIVAKGEHKSGVLDEPTIATILREVLEGLEYLHKNGQIH
SPAK_Human    MLDIIKYIVNRGEHKNGVLEEAIIATILKEVLEGLDYLHRNGQIH
SPAK_Mouse    MLDIIKYIVNRGEHKNGVLEEAIIATILKEVLEGLDYLHRNGQIH
SPAK_Rat      MLDIIKYIVNRGEHKNGVLEEAIIATILKEVLEGLDYLHRNGQIH
FRAY_Drosop   LLDIIKHKMRTSNCKQGVFDEATIATVLKEVLKGLEYFHSNGQIH

cons          :*****: :   .: *.**::*. ***:*:***:**:*:* *****


OSR1_Frog     RDVKAGNILLGDDGSVQIADFGVSAFLATGGDITRNKVRKTFVGT
OSR1_Human    RDVKAGNILLGEDGSVQIADFGVSAFLATGGDITRNKVRKTFVGT
OSR1_Mouse    RDVKAGNILLGEDGSVQIADFGVSAFLATGGDITRNKVRKTFVGT
SPAK_Human    RDLKAGNILLGEDGSVQIADFGVSAFLATGGDVTRNKVRKTFVGT
SPAK_Mouse    RDLKAGNILLGEDGSVQIADFGVSAFLATGGDVTRNKVRKTFVGT
SPAK_Rat      RDLKAGNILLGEDGSVQIADFGVSAFLATGGDVTRNKVRKTFVGT
FRAY_Droso    RDIKAGNILIGDDGTIQIADFGVSAWLATGRDLSRQKVRHTFVGT

cons          **:*******:*:**::**********:**** *::*:***:****
```

```
OSR1_Frog    PCWMAPEVMEQVRGYDFKADIWSFGITAIELATGAAPYHKYPPMK
OSR1_Human   PCWMAPEVMEQVRGYDFKADIWSFGITAIELATGAAPYHKYPPMK
OSR1_Mouse   PCWMAPEVMEQVRGYDFKADIWSFGITAIELATGAAPYHKYPPMK
SPAK_Human   PCWMAPEVMEQVRGYDFKADMWSFGITAIELATGAAPYHKYPPMK
SPAK_Mouse   PCWMAPEVMEQVRGYDFKADMWSFGITAIELATGAAPYHKYPPMK
SPAK_Rat     PCWMAPEVMEQVRGYDFKADMWSFGITAIELATGAAPYHKYPPMK
FRAY_Droso   PCWMAPEVMEQDHGYDFKADIWSFGITAIEMATGTAPYHKYPPMK

cons         *********** :*******:*********:***:*********


OSR1_Frog    VLMLTLQNDPPTLETGVQDKEMLKKYGKSFRKMISSCLQKDPEKR
OSR1_Human   VLMLTLQNDPPSLETGVQDKEMLKKYGKSFRKMISLCLQKDPEKR
OSR1_Mouse   VLMLTLQNDPPSLETGVQDKEMLKKYGKSFRKMISLCLQKDPEKR
SPAK_Human   VLMLTLQNDPPTLETGVEDKEMMKKYGKSFRKLLSLCLQKDPSKR
SPAK_Mouse   VLMLTLQNDPPTLETGVEDKEMMKKYGKSFRKLLSLCLQKDPSKR
SPAK_Rat     VLMLTLQNDPPTLETGVEDKEMMKKYGKSFRKLLSLCLQKDPSKR
FRAY_Droso   VLMLTLQNDPPTLDTGADDKDQYKAYGKTFRKMIVECLQKEPSKR

cons         *************:*:**.:**:   * ***:***::   ****:*.**


OSR1_Frog    PTAAELLKHKFFQKSKNK
OSR1_Human   PTAAELLRHKFFQKAK--
OSR1_Mouse   PTAAELLRHKFFQKAK--
SPAK_Human   PTAAELLKCKFFQKAK--
SPAK_Mouse   PTAAELLKCKFFQKAK--
SPAK_Rat     PTAAELLKCKFFQKAK--
FRAY_Droso   PTASELLKHAFFKKAK--

cons         ***:***:  **:*:*
```

**Table 4-1.  Sequence Alignment of GCK-IV Subfamily Protein Kinase**
The sequence alignment was generated by program *T-Coffee*.  Among the sequence aligned, only OSR1 kinase domain structure was determined.  These OSR1/SPAK (Ste20 Proline Alanine rich Kinase) homologs of kinase domain reveal more than 90% sequence homology.

The signature sequence (GTPCWMAPEV) is almost overlapped with the typical P+1

loop sequence, which is thought to play an integral role in substrate recognition.  Since

this signature sequence of Ste20 kinase family is P+1 loop, it can be proposed that this kinase family might share similar substrate specificity.

**General Topology**

The crystal structure of human *Ste20* OSR1 kinase domain reveals a domain-swapped dimeric formation as shown in **Figure 4-1 and 4-2**, although it was monomeric in solution. Each subunit of dimeric structure stands up in face-to-face vertical manner, but each monomer is tilted away from each other by about 40º as shown in **Figure 4-3**. In terms of general topology, this structure presents several lines of unique features, which were not observed previously in the study of protein kinase structures.

First, the crystal structure of human *Ste20* OSR1 kinase domain reveals a novel mode of homo-dimerization of protein kinase. Although it is not uncommon that protein kinase is dimerized, this OSR1 kinase structure is unique in that C-terminal lobe is mainly involved in this protein-protein interaction, not N-terminal lobe. In general, however, the common strategies adopted by protein kinase for dimerization utilize N-terminal lobe as interacting motif. In OSR1 kinase structure, N-terminal lobe from each subunit is intact from this novel mode of dimerization, in which each monomer stands face to face.

Secondly, each monomer of this novel dimeric formation reveals the highly conserved kinase fold, except for the fact that part of c-terminal lobe was apparently "domain-swapped" with the other subunit. As mentioned earlier, each monomeric OSR1 kinase structure shows typical characteristics of inactive protein kinase, since this kinase construct (1~295) is missing in PF-1 domain, which is required for kinase activity.
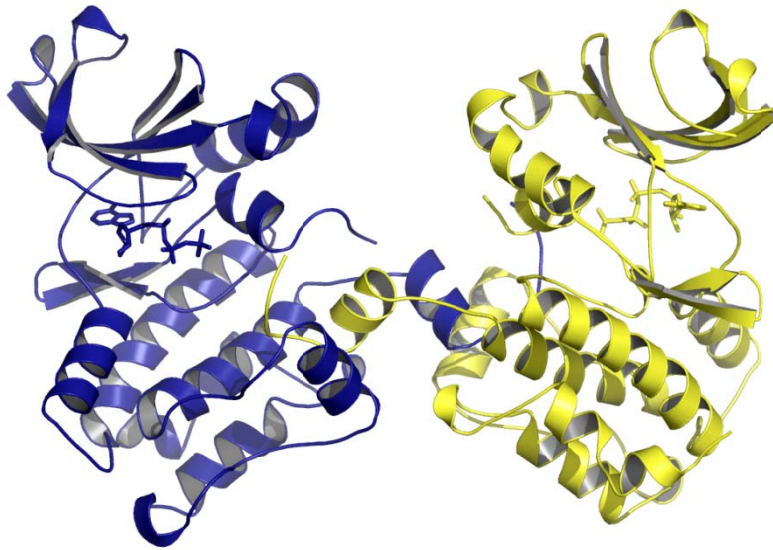
**Figure 4-1.  Dimeric Structure of OSR1 Kinase Domain (front view)**
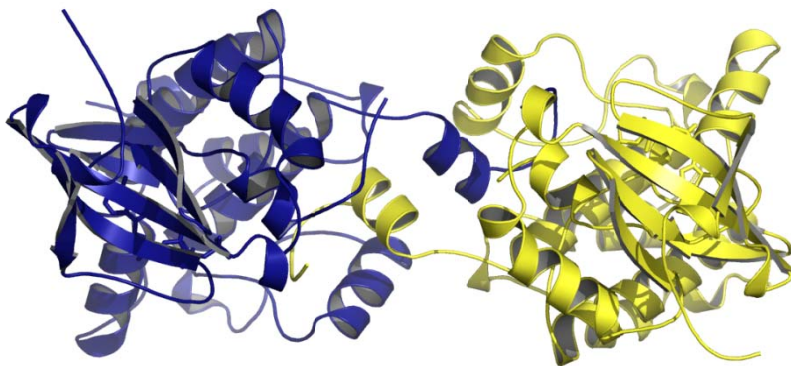


**Figure 4-2.  Dimeric Structure of OSR1 Kinase Domain (bottom view)**
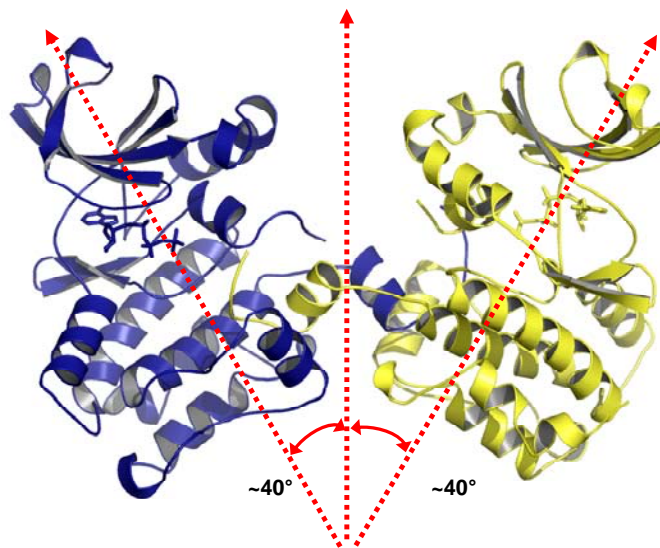
**Figure 4-3.  The tilted kinase domain relative to the other subunit**
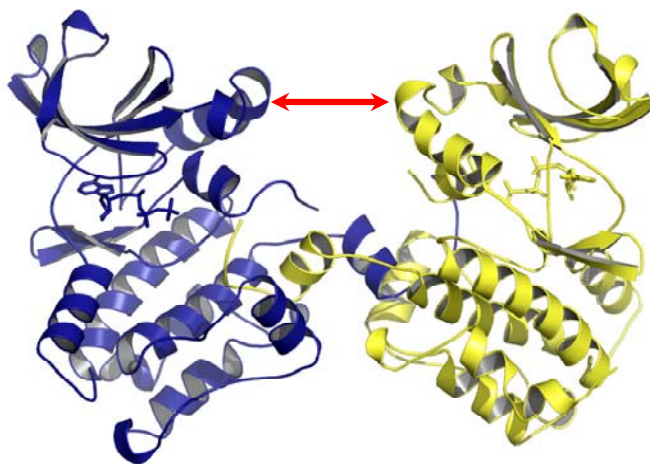


**Figure 4-4.   Relative physical arrangement of N-terminal lobe in dimeric OSR1 structure**

**Domain Swapping in the OSR1 kinase domain**

Although it is clear that dimeric structure of OSR1 kinase is domain-swapped in terms of general topology, it should be investigated whether the loop region termed as swapped is really "swapped" or "penetrated". Since domain swapping involves the exchange of functional unit, most of chemical interactions must be intact compared to "prior-to-swapping" state, especially in terms of overall folding. If a given portion of protein undergoes significant conformational change or disruption following so-called "domain-swapping", such a rearrangement should be considered partial unfolding or penetration, not domain-swapping. The best strategy to differentiate between bona-fide domain swapping and blunt penetration related to partial unfolding is to superimpose the protein structure of interest with intact homology model for a detailed comparison.

As shown in **Figure 4-5**, another *Ste20* kinase TAO2 (PDB: 1U5R) was aligned with OSR1 kinase domain. In addition to general alignment over the entire kinase domain, P+1 loop region was well aligned with the corresponding portion of TAO2 kinase domain. In the **Figure 4-5**, the only backbone of kinase domain was presented to maximize the clarity of general topology. The helical turns from *αEF* helix beyond the P+1 loop was completely aligned. Based upon the structural alignment between TAO2 and OSR1 kinase, it is clear that domain-swapped region corresponds to the region including P+1 loop and the following *αEF* helix.
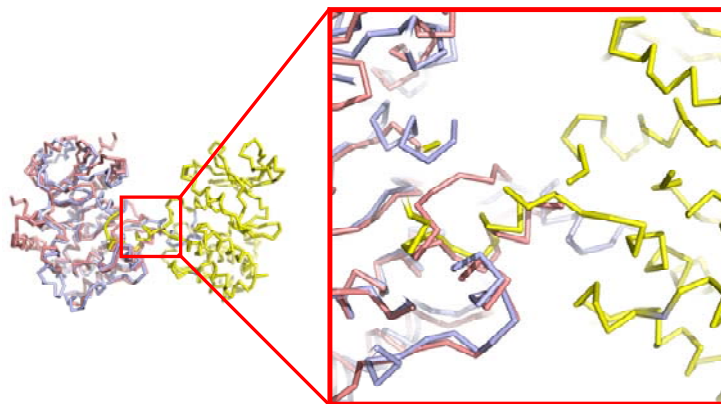
**Figure 4-5.  Structural Alignment of OSR1 Kinase Domain with TAO2**
Blue and yellow models represent the dimeric structure of OSR1 kinase domain, whereas magenta model indicates TAO2 kinase domain.
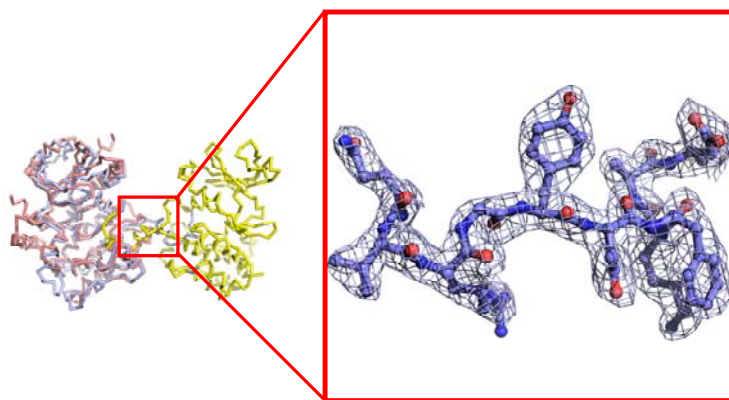


**Figure 4-6. Electron Density map from mid-zone linking each subunit**
The electron density map from linking region between two subunits of dimer is clearly represented.  In the red box, Val 201, Arg 202, Gly 203, Tyr 204, Asp 205, Phe 206 (from left to right) are presented with electron density map ($2F_o$- $F_c$).

Because the structure of OSR1 kinase domain was solved by molecular replacement, there might be some chance of model bias associated with its topology. As the free $R$ value of the OSR1 kinase structure refined to 2.25 Å, however, is about 26.7%, there is no theoretical room for model bias. Furthermore, the electron density map ($2F_o$-$F_c$) correctly reveals the region between two subunits as shown in **Figure 4-6**. Thus, it is now obvious that the dimeric crystal structure of OSR1 kinase underwent domain-swapping via the middle of polypeptide chain termed as activation segment.

From the viewpoint of structural biology, domain-swapped protein kinase is quite surprising. Since the crystal structure of PKA (Protein Kinase A) was reported in 1991, there has been no case of domain-swapping via the activation segment in the protein kinase family, except for atypical protein kinase domain of TRP channel (Yamaguchi, Matsushita et al. 2001), which was domain-swapped through the N-terminal helical extension. This atypical kinase structure of TRP channel was covered in detail in the later section.

According to our current understanding of domain-swapping, such an event is mainly facilitated by the partially unfolded N-terminal or C-terminal region, which is coupled to the hinge loop and then stable anchor portion (Liu and Eisenberg 2002). In many cases, domain-swapping is coupled to mechanism of protein aggregation and fibril (amyloid) formation associated with deposition diseases (Bennett, Sawaya et al. 2006).

Since most of domain-swapped protein structures are involved in irreversible deposit formation leading to dead-lock conformation, it is quite intriguing and unexpected that the protein kinase actively utilized in signal transduction via its reversible conformational change can be actually domain-swapped.

Surprisingly, however, right after the determination of OSR1 kinase structure, the extensive PDB database search found several more cases of domain-swapping in the protein kinase family. In addition to OSR1, the other protein kinases that were shown to be domain swapped include SLK (Ste20-Like Kinase), STK10 (Ser/Thr Kinase 10), and DAPK3 (Death Associated Protein Kinase 3). Among these domain-swapped protein kinases, SLK and STK10 belong to *Ste20* kinase family. Especially, it was revealed that SLK was domain-swapped in non-, mono-, di-phosphorylated form. In all the cases of SLK, STK10, and DAPK3, the same region including P+1 loop and *αEF* helix were domain-swapped in a similar manner to OSR1. The detailed discussions on those structures were given in later section.

In order to compare the basic topology of each domain-swapped protein kinases with PKA, those structure were aligned based upon the backbone model as shown in **Figure 4-7**. As expected, general topology was well aligned with each other, except for activation segment. The catalytiv loop as well as DFG region (magnesium binding loop) was well aligned, together with *αF* helix.

As shown in **Figure 4-8**, it is shown that the sequence after magnesium bind loop (DFG) begins to be perturbed. In the same fashion, the sequence prior to F-helix is also perturbed. Thus, it can be suggested that DFG magnesium binding loop and N-terminal F-helix seem to function as anchor for hinge loop, based upon the structural alignment with other domain-swapped Ste20p kinases, which were reported after OSR1 structure determination.

In addition, the multiple sequence alignment analysis between domain-swapped kinases suggests that the conserved sequence (**R-x-x-F**) located in the c-terminal region

of activation loop might play a certain role, which remains unclear thus far as shown in

**Table 4-2**. The sequence is uniquely conserved only among the domain-swapped protein

kinases. Another interesting observation is from polar interaction between catalytic

lysine and tryptophan from P+1 loop, because this intermolecular interaction in the

swapped region is well conserved among the domain-swapped protein kinases from

Ste20 family, as indicated in **Figure 4-9**. This polar interaction between swapped

polypeptide is characterized by cation-π interaction, which is frequently found in other

non-kiase domain-swapped protein structures. However, as this interaction is also found

in the same location from non-swapped Ste20 PAK1 or TAO2, it is yet to be investigated

how this cation-π interaction contributes to domain-swapping in the protein kinase.
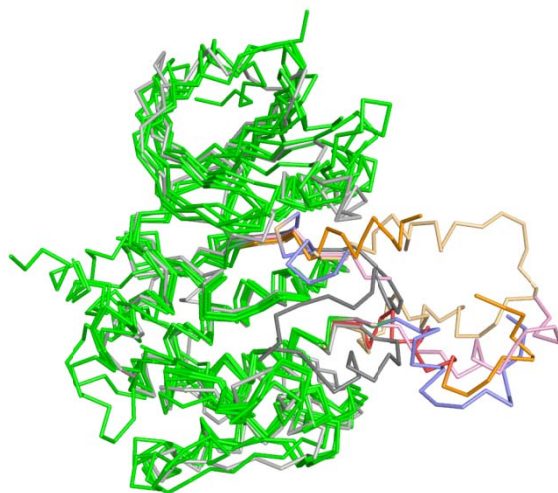


**Figure 4-7. Structural Alignment of OSR1 kinase with other Domain-Swapped protein kinase and PKA**
Gray indicates PKA (as standard model for comparison), all other structures were represented by green, except for activation segment (from DFG ~ to APE). Orange indicates STK10, blue indicates OSR1, bright orange indicates non-phosphorylated SLK, and magenta indicates DAPK3.
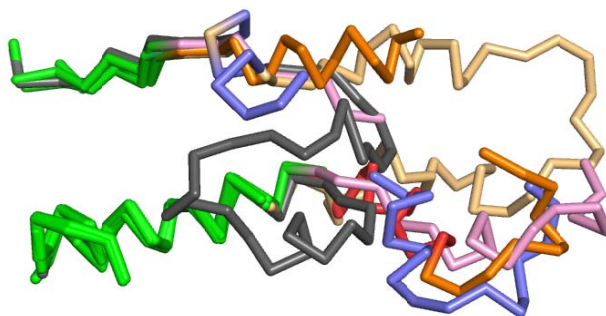
**Figure 4-8. Structural Alignment of Activation Segments from PKA and Domain-Swapped protein kinases.**

Gray indicates PKA (as standard model for comparison), orange indicates STK10, blue indicates OSR1, bright orange indicates non-phosphorylated SLK, and magenta indicates DAPK3.
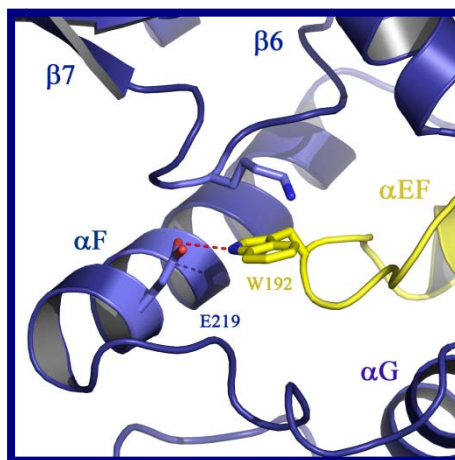


**Figure 4-9. Polar Interaction between domain-swapped subunits**

Bule and yellow model indicate each polypeptide. The polar interaction called cation-π interaction is clearly found between lysine from catalytic loop and tryptophan from P+1 loop. Also, hydrogen bond between E219 and W192 is visualized. Except for these polar interactions, most of the residues in the interface inteact with each other via hydrophobic interaction.

```
2J7T    PPHHELNPMRVLLKIAKSDPPTLLTP-------SKWSVEFRDFLKIALDKNPET
2JFL    PPHHELNPMRVLLKIAKSEPPTLAQP-------SRWSSNFKDFLKKCLEKNVDA
OSR1    APYHKYPPMKVLMLTLQNDPPSLETGVQDKEMLKKYGKSFRKMISLCLQKDPEK
1ATP    PPFFADQPIQIYEKIVSGK---VRFP-------SHFSSDLKDLLRNLLQVDLTK
1F3M    PPYLNENPLRALYLIATNGTPELQNP-------EKLSAIFRDFLNRCLDMDVEK
1U5R    PPLFNMNAMSALYHIAQNESPALQS--------GHWSEYFRNFVDSCLQKIPQD

cons    .*      .:        .      :          :  .  ::.::   *:


2J7T    RP-----SAAQLLEHPFVSSITSNKAL--------------------------
2JFL    RW-----TTSQLLQHPFVT-VDSNKPI--------------------------
OSR1    RP-----TAAELLRHKFFQKA--------------------------------
1ATP    RFGNLKNGVNDIKNHKWFATTDWIAIYQRKVEAPFIPKFKGPGDTSNFDDYEEE
1F3M    RG-----SAKELLQHQFLKIAKPLSSL--------------------------
1U5R    RP-----TSEVLLKHRFVLRER-------------------------------

cons    *         :  .* :.


2J7T    --RELVAE--AKAEVMEE------
2JFL    --RELIAE--AKAEVTEEVEDGKE
OSR1    ----------------------K
1ATP    EIRVSINEKCGKEFTEF-------
1F3M    --TPLIAA--AKEATKNN-----H
1U5R    ----------------------

cons
```

**Table 4-2. Sequence Alignment of Domain Swapped Protein Kinases and Non-Swapped Protein Kinses.**

The sequence alignment was generated by program *T-Coffee*.

2J7T indicates STK10 (Serine/Threonine Kinase 10), 2JFL represents SLK (Ste20-Like Kinase), 1ATP indicates PKA (Protein Kinase A), 1F3M indicates PAK1 and 1U5R indicates Ste20 TAO2.

**Domain Swapping in the SLK (S̲te20-L̲ike K̲inase)**

The Ste20 kinase SLK (S̲terile20-L̲ike K̲inase) is a ser/thr kinase and belongs to GCK (G̲erminal C̲enter K̲inase) V subfamily, which also includes LOK (L̲ymphocyte O̲riented K̲inase) (Ellinger-Ziegelbauer, Karasuyama et al. 2000). Like other GCK subfamily, SLK contains N-terminal kinase domain and C-terminal regulatory domain.

Human SLK kinase domain shares more than 70% homology with human OSR1 kinase domain whose crystal structure was shown to be domain-swapped dimer. Following the determination of OSR1 crystal structure, the structure of human SLK was reported to Protein Data Bank database without published discussion. Surprisingly, it is revealed that human SLK forms domain-swapped dimer in a strikingly similar manner to OSR1 kinase domain structure as shown in **Figure 4-10** and **4-11**. Each subunit from dimeric structure interacts with the other subunit via c-terminal-lobe-involving swapping strategy, which is also shown in OSR1 kinase structure. As discussed with OSR1 kinase structure earlier, activation segment c-terminal fragment called P+1 loop was swapped and located in the other subunit like its own folded protein. This SLK crystal structure provides clear evidence that OSR1 kinase structure is not a byproduct of crystallization artifact.

Moreover, additional crystal structure from mono- and diphosphorylated human SLK turned out to share almost identical topology with non-phosphorylated SLK as shown in Figure 4-12~17. As of now, since the mechanism of SLK regulation is still elusive, it is difficult to address how this domain-swapping is involved in kinase activity regulation and also how fully phosphorylated kinase can still take inactive conformation with regard to substrate binding interaction. From the structural viewpoint, as even di-

phosphorylated SLK forms a swapped-dimer and provide no room for substrate binding, it seems very likely that there might be another layer of structural regulation event, which triggers substantial conformational changes that allow this SLK kinase to have access to substrate or to release each subunit from swapped dimer. Thus, it follows that we need to understand more biological context in terms of SLK regulation, which might help explain why domain swapping is utilized and what kind of advantages the Ste20p family benefits from, compared to other known mechanisms to regulate signal transduction in the protein kinase.

Recently, it is reported that SLK is involved in actin stress fiber dissolution and mediation of apoptosis (Storbeck, Daniel et al. 2004). Also, it is shown that SLK is necessary for cell cycle progression through G2 phase and also associated with muscle development as well as with adult skeletal muscle (Ellinger-Ziegelbauer, Karasuyama et al. 2000; Wagner, Flood et al. 2002).

SLK is known to be highly expressed in the muscle mass and neuronal lineages of developing embryos (Wagner, Flood et al. 2002). Expression of SLK is detected in the forebrain, midbrain and hindbrain of the developing CNS. Also, it is reported that SLK is expressed in the hypothalamus region, all layers of the neural tube, dorsal root ganglion and in the proliferating ependymal layers (Storbeck, Daniel et al. 2004).

More interestingly, it is shown that SLK is cleaved by caspase 3 in vitro and in vivo during c-Myc-, TNF (tumor necrosis factor) α and UV-induced apoptosis (Sabourin and Rudnicki 1999). Cleavage of SLK by caspase 3 generates two domains, which have distinct enzyme activities; an activated N-terminal kinase domain which induces apoptosis and cytoskeletal rearrangements and a C-terminal domain which disassembles

actin stress fibers.  It is also known that SLK can interact with neuronal transmembrane

protein calsyntenin and activate Polo-like kinase 1 (Wagner, Flood et al. 2002).
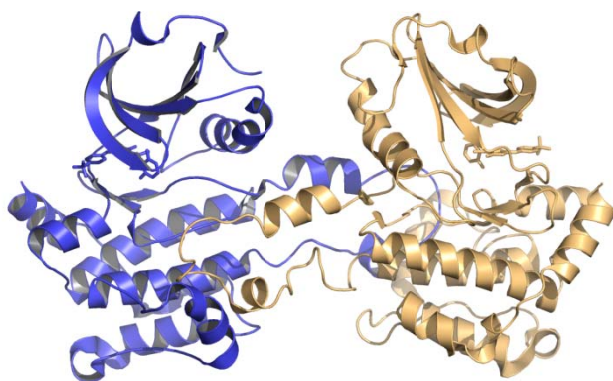


**Figure 4-10.   Front view of the crystal structure of Non-phosphorylated SLK (Sterile20-Like Kinase) PDB code: 2J51**
This figure shows domain-swapped dimeric structure of non-phosphorylated human SLK (Sterile20-Like Kinase), which adopts a strikingly similar mode of dimerization to OSR1 kinase domain.  Compared to OSR1 kinase domain-swapped dimer, each monomer stands up in a more parallel manner. Blue and light orange models represent each subunit.
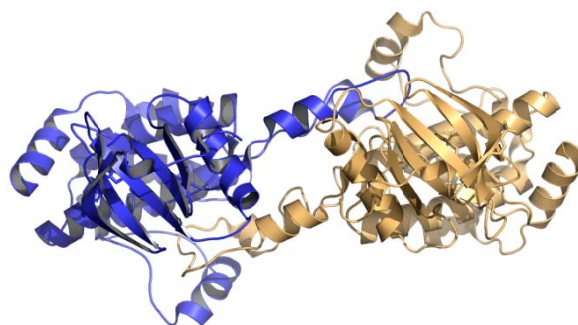


**Figure 4-11.   Bottom view of the crystal structure of Non-phosphorylated SLK (Sterile20-Like Kinase) PDB code: 2J51**
This view of the domain-swapped crystal structure of SLK was taken from the bottom in order to help show the overall topology of dimerization, which is also adopted in OSR1 kinase structure.   Blue and light orange models represent each subunit.
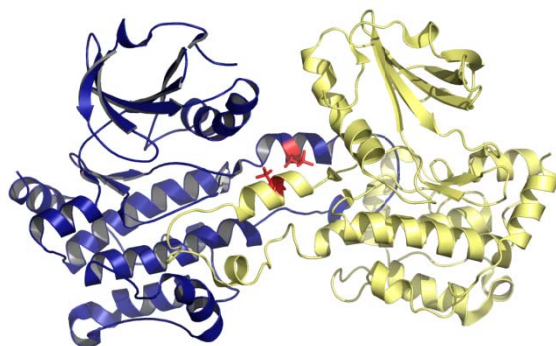
**Figure 4-12.  Front view of the crystal structure of Mono-phosphorylated human SLK (Sterile20-Like Kinase) PDB code:2JFM**

This figure shows domain-swapped dimeric structure of mono-phosphorylated human SLK (Sterile20-Like Kinase), which adopts a strikingly similar mode of dimerization to OSR1 kinase domain and also to non-phosphorylated swapped dimer (PDB code: 2J51). Blue and light yellow models represent each subunit.

The red residue represents phosphorylated residue.
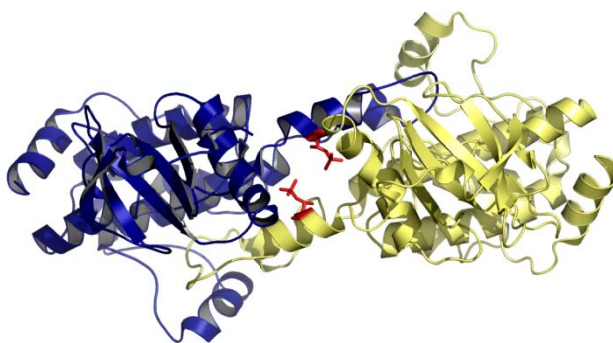


**Figure 4-13. Bottom view of the crystal structure of mono-phosphorylated human SLK (Sterile20-Like Kinase) PDB: 2JFM**

This view of the domain-swapped crystal structure of SLK was taken from the bottom in order to help show the overall topology of dimerization, which is also adopted in OSR1 kinase structure.  Blue and light yellow models represent each subunit.  The red residue represents phosphorylated residue.
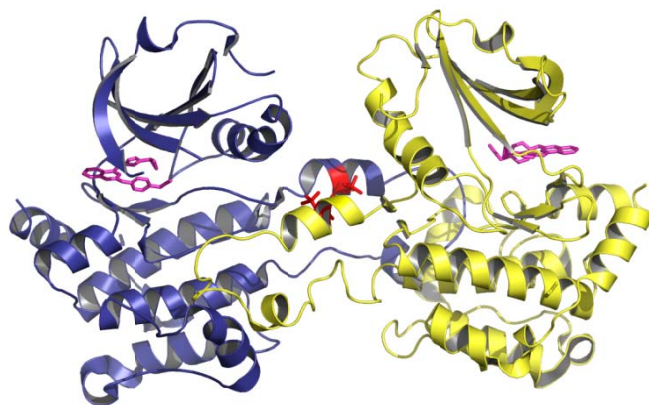
**Figure 4-14. Front view of the crystal structure of mono-phosphorylated human SLK (<u>S</u>terile20-<u>L</u>ike <u>K</u>inase) complexed with K00606a**
**PDB code: 2UV2**
This figure shows domain-swapped dimeric structure of mono-phosphorylated human SLK (<u>S</u>terile20-<u>L</u>ike <u>K</u>inase), which adopts a strikingly similar mode of dimerization to OSR1 kinase domain and also to non-phosphorylated swapped dimer (PDB code: 2J51). Blue and light yellow models represent each subunit. The red residue represents phosphorylated residue. Magenta represents inhibitor K00606a.
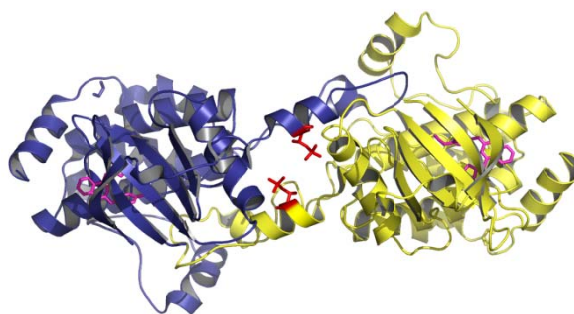


**Figure 4-15. Bottom view of crystal structure of mono-phosphorylated human SLK (<u>S</u>terile20-<u>L</u>ike <u>K</u>inase) complexed with K00606a. PDB: 2UV2**
This view of the domain-swapped crystal structure of SLK was taken from the bottom in order to help show the overall topology of dimerization, which is also adopted in OSR1 kinase structure and also non-phosphorylated swapped dimer (PDB code: 2J51). Blue and light yellow models represent each subunit. The red residue represents phosphorylated residue. Magenta represents inhibitor K00606a.
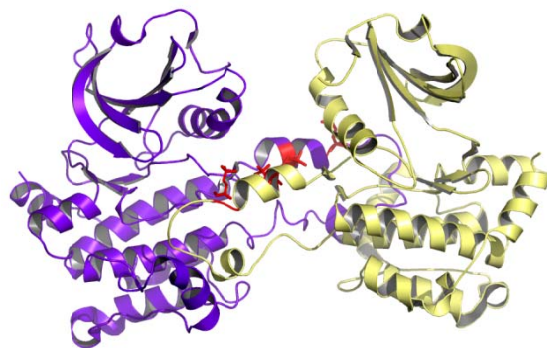
**Figure 4-16.  Front view of the crystal structure of di-phosphorylated human SLK (Sterile20-Like Kinase) PDB code: 2JFL**
This figure shows domain-swapped dimeric structure of di-phosphorylated human SLK (Sterile20-Like Kinase), which adopts a strikingly similar mode of dimerization to OSR1 kinase domain and also to non-, mono-phosphorylated swapped dimer (PDB code: 2J51 and 2UV2).  Blue and light yellow models represent each subunit.  The red residue represents phosphorylated residue.
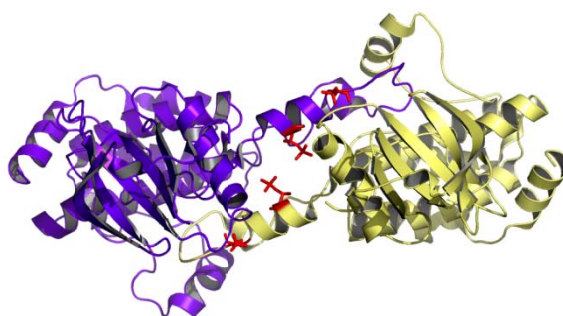


**Figure 4-17.  Bottom view of the crystal structure of di-phosphorylated human SLK (Sterile20-Like Kinase) PDB: 2JFL**
This view of the domain-swapped crystal structure of di-phosphorylated SLK was taken from the bottom in order to help show the overall topology of dimerization, which is also adopted in OSR1 kinase structure and other non- and mono-phosphoryalted dimeric structures.  Blue and light yellow models represent each subunit.  The red residues represent phosphorylated ones.

**Domain Swapping in the STK10 (Ser/Thr Kinase 10)**

STK10 (Ser/Thr Kinase 10) is another interesting protein kinase which was shown to form a domain-swapped dimer in the crystal structure as shown in **Figure 4-18** and **4-19**. Like the previously discussed structure such as OSR1 and SLK, P+1 loop portion from activation segment was swapped and well-folded. As in the case of SLK, only structure information from PDB (Protein Data Bank) is available, without any published discussions about STK10 crystal structure. However, the domain-swapped STK10 crystal structure adds additional evidence of domain swapping in the protein kinase. Like OSR1 as well as SLK, STK10 belongs to Sterile 20 kinase family, also human homolog of the murine kinase LOK (Lymphocyte-Oriented Kinase) (Walter, Cutler et al. 2003). The crystal structure of LOK is not available now, although LOK shares more than 80% homology with STK10.

It is known that STK10 is closely related to the polo-like kinase kinase SLK and reported to associate with Plk1 and activate it *in vitro* (Kuramochi, Moriguchi et al. 1997). The high STK10 expression is detected in certain proliferating tissues (spleen, placenta, and peripheral blood leukocytes), but transcripts were also detectable in other tissues profiled (brain, heart, skeletal muscle, colon, thymus, kidney, liver, small intestine, and lung) indicating that STK10 has a more widespread pattern of expression than mouse ortholog LOK (Kuramochi, Matsuda et al. 1999). Also, STK10 has been found to be overexpressed in a variety of tumor cells and a heterozygous somatic missense mutation (K277E) within the kinase domain has been reported in human testicular germ-cell tumors.
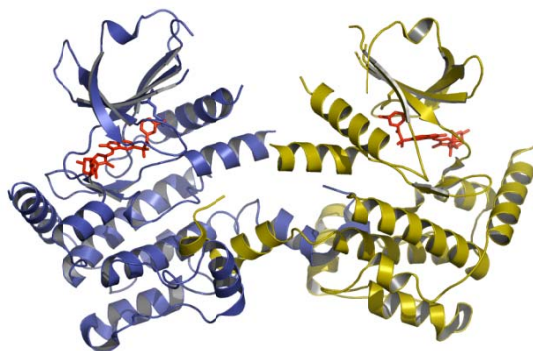
**Figure 4-18. Front view of the crystal structure of non-phosphorylated human STK10 (Serine Threonine Kinase 10) complexed with pyrrole indolinone SU11274. PDB code: 2J7T**
This figure shows domain-swapped dimeric structure of non-phosphorylated human STK10 (Serine Threonine Kinase 10), which adopts a strikingly similar mode of dimerization to OSR1 kinase domain and also to non-, mono-and di-phosphorylated swapped SLK dimer. Blue and dark yellow models represent each subunit. The red molecule represents cell-permeable pyrrole indolinone SU11274.
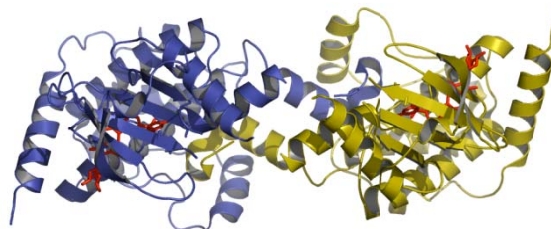


**Figure 4-19. Bottom view of the crystal structure of non-phosphorylated human STK10 (Serine Threonine Kinase 10) complexed with pyrrole indolinone SU11274. PDB code: 2J7T**
This view of the domain-swapped crystal structure of SLK was taken from the bottom in order to help show the overall topology of dimerization, which is also adopted in OSR1 kinase structure and domain-swapped SLK structures. Blue and dark yellow models represent each subunit. The red molecule represents cell-permeable pyrrole indolinone SU11274.

**Domain Swapping in the DAPK3 (<u>D</u>eath-<u>A</u>ssociated <u>P</u>rotein <u>K</u>inase <u>3</u>)**

An extensive PDB database search reveals another case of domain-swapped protein kinase, human DAPK3. Human DAPK3 crystal structure was reported to form typical domain-swapped dimerization as shown in OSR1 or SLK as shown in **Figure 4-20** and **4-21**. Although sequence homology between DAPK and Ste20p kinase is approximately 60%, overall topology of domain-swapped dimer is basically identical to OSR1 or SLK. C-terminal activation segment was mainly involved in domain-swapping via P+1 loop region.

DAPK, a ser/thr protein kinase, was identified in an unbiased genetic screen aimed at finding positive mediators of cell death (Shani, Marash et al. 2004). It has been characterized as a tumor suppressor protein whose expression is greatly reduced or lost in a number of human cancers. Several protein kinases such as DAPK (Death-Associated Protein Kinase), DAP kinase-related protein 1, and DAPK3 (also called ZIPK:Zipper-Interacting Protein Kinase) belong to DAP Kinase family. DAPK belongs to the family of $Ca^{2+}$-CAM-regulated kinases and contains a $Ca^{2+}$-CAM binding domain proximal to the catalytic domain (Shani, Marash et al. 2004). The multidomain structure of DAPK also includes a stretch of ankyrin repeats, acytoskeleton binding domain and a death domain shown to be required in some of the DAPK-induced cell death scenarios.

Members of this DAPK family are known to mediate cell death triggered by various external and internal signals such as TNF-α, FAS and transforming growth factor β and detachment from the extracellular matrix. Especially, DAPK3 kinase localizes to both the nucleus and the cytoplasm and fractionates as monomeric and trimeric forms. DPAK family members are $Ca^{2+}$-CAM-regulated kinases associated with actin

microfilaments (Page, Kogel et al. 1999).  However, DAPK3 lacks the CaM regulatory

domain, indicating different modes of activation.  In addition to the N-terminal catalytic

domain, DAPK3 contains a leucine zipper structure at its C-terminus which is thought to

be mediate homodimerization (Page, Kogel et al. 1999).
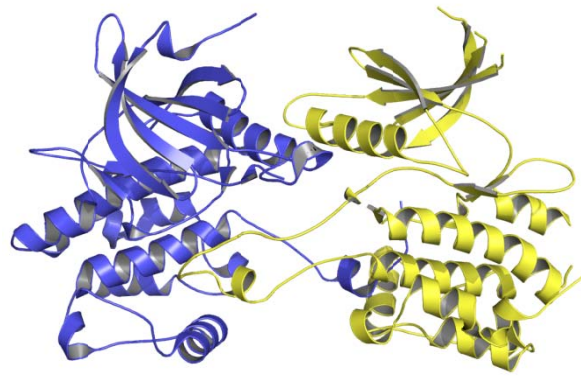


**Figure 4-20.  Front view of the crystal structure of non-phosphorylated human DAPK3 (Death Associated Protein Kinase 3). PDB code: 2J90**
This figure shows domain-swapped dimeric structure of non-phosphorylated human DAPK3 (Death Associated Protein Kinase 3), which adopts a generally similar mode of dimerization to OSR1 kinase domain and also to non-, mono-and di-phosphorylated swapped SLK dimers.  However, intermolecular reaction via domain-swapping doesn't seem to be so tight as other reported swapped structures.  Blue and dark yellow models represent each subunit.
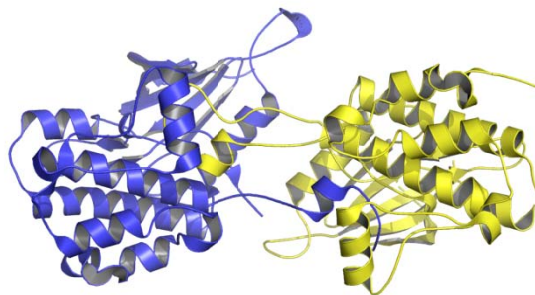
**Figure 4-21. Bottom view of the crystal structure of non-phosphorylated human DAPK3 (D̲eath A̲ssociated P̲rotein K̲inase 3). PDB code: 2J90**

This view of the domain-swapped non-phosphorylated human DAPK3 (D̲eath A̲ssociated P̲rotein K̲inase 3) was taken from the bottom in order to help show the overall topology of dimerization, which reveals similar but more loose intermolecular interaction, compared to OSR1 kinase structure and domain-swapped SLK structures. Blue and dark yellow models represent each subunit.

**Domain Swapping in the atypical protein kinase**

Another interesting case of domain swapping in the protein kinase is atypical protein kinase domain of TRP (T̲ransient R̲eceptor P̲otential) channel as shown in **Figure 4-22** (Yamaguchi, Matsushita et al. 2001). As shown in Figure, each monomer interacts with each other via N-terminal lobe by an extended helical structure. There is no inter-C-terminal lobal interaction, which was common in the previous cases like SLK and OSR1. Unlike the traditional TRP channel proteins, this novel TRP channel protein has ability to phosphorylate itself and other proteins on serine and threonine residues. Since the catalytic domain of this novel TRP channel protein has no clear sequence homology to classical eukaryotic protein kinases, this kinase domain is classified as atypical protein kinase. However, this (atypical) kinase domain of TRP channel reveals unexpected similarity to classical eukaryotic protein kinase. Given the current understanding of

dimeric atypical protein kinase of TRP, it is not clear why this domain-swapping is employed and how this contributes to signal transduction.

TRP channels are known to modulate calcium levels in eukaryotic cells in response to external signals (Yamaguchi, Matsushita et al. 2001). TRP channels are integral membrane proteins in which the ion-conducting pores are formed by six membrane-spanning helical segments that are similar to those of voltage-gated potassium channels and cyclic nucleotide-gated channels. Also, the combination of a $Ca^{2+}$ channel and a protein kinase domain within a single polypeptide chanin is rare and unusual (Yamaguchi, Matsushita et al. 2001). Although no other known ion channel possesses a kinase domain within the same polypeptide chain, many channels are known to be regulated by phosphorylation. Like other domain-swapped protein kinaes, there is no clue why this swapping is necessary for the regulation of TRP channel and how this swapping is controlled.

**Figure 4-22. Crystal Structure of Atypical TRP Channel Kinase Domain PDB: 1IAJ**
This figure shows domain-swapped dimeric structure of non-phosphorylated atypical kinase domain of TRP channel proteion, which adopts a completely different mode of dimerization from OSR1 kinase domain and other domain-swapped SLK dimers. Blue and light yellow models represent each subunit.

**Introduction: 3-Dimensional Domain Swapping**

*General Introduction*

Domain swapping is defined as the mechanism by which two or more protein molecules exchange part of their structures to form intertwined oligomers (Rousseau, Schymkowitz et al. 2003). When domain swapping takes place, two or more subunits exchange identical structural elements or "domains". These so-called "domains", however, implies a wide range of structural units in terms of the extent of swapped protein subunit. The term "domain" could indicate either simple secondary structure elements or complete, globular, functional domain. In the case of domain-swapped oligomers, a given structural unit of one subunit replaces the identical structural unit of another subunit, giving rise to the reorganization of the monomeric fold contributed from different subunits.

If both the monomer and dimer of a molecule exist in stable forms, in which the dimer adopts a domain-swapped conformation and the monomer adopts a closed conformation, this protein is regarded as a bona fide example of domain swapping (Liu and Eisenberg 2002). Some proteins, however, form interwined, domain-swapped oligomers without a known closed monomer. If these proteins have homologs known to be closed monomers, these oligomers are classified into "quasidomain-swapped" (Liu and Eisenberg 2002). If a protein forms an oligomer by exchanging domains, but there is no monomeric form or homolog for the protein, this protein is considered a candidate for 3D domain swapping.

*Examples of domain swapping in other proteins*

I. *Bona fide* domain swapping

• diphtheria toxin

• RNases (A, BS-)

• CksHs (cell cycle regulation)

• CD2 (T-cell adhesion)

• staphylococcal nuclease

• λ-cro (DNA repressor)

• α-spectrin (cytoskeleton)

• antibody fragments

• NO reductase

• NOS(ox)  (redox protein)

• Spo0A (development regulation)


II.  *Quasi* domain swapping

• crystallin

• growth factors/cytokines

• pheromone/odorant binding/transport proteins

• SH3 (signal transduction)

• glyoxalase I

• RYMV (viral capsid protein)

• human cystain C (protease inhibitor)

*Advances in 3D domain swapping*

Thus far, about more than 40 domain-swapped molecules have been reported (Jaskolski 2001). One common aspect of those domain-swapped molecules is the location of "swapped" region in the original subunit. Most of those "domain-swapped" regions are from N-terminus or C-terminus. It is no wonder to suspect that these extreme ends of any protein might have higher propensity for "unfolding" process, which may be associated with rearrangement events including "domain-swapping", because it is assumed that these terminal regions are likely to be intrinsically flexible. Furthermore, we can also assume that such a partial unfolding event associated with "domain-swapping" could propagate into more distant region. In several cases such as $\beta$B2-crystallin, cyanovirin-N and calbindin $D_{9k}$, half of the molecule is swapped (Jaskolski 2001).

In some cases, more than one domain in a protein can swap. Recently, it was shown that both N-termini and C-termini were domain-swapped in RNase A dimer. Also, it has been shown that RNase A can form oligomers like trimer or tetramer. This observation suggests that there might exist more than one mode of domain-swapping in a given protein. The existence of diverse modes of domain-swapping in the same protein tells us the possibility that this domain-swapping event is not just the product of intrinsically loaded mechanism towards stabilization, also the result of intermolecular interactions activated by external stimuli, depending on physiological environment. It thus follows that domain-swapping actually opens a window of research opportunities towards understanding various modes of intra- or inter- molecular interaction, which might help explain bigger picture of protein folding.

*The swapped domains are diverse*

The swapped domains turned out to contain various sizes as well as sequences, in addition to its secondary structure (Bennett and Eisenberg 2004). The swapped region might be entire functional domain made of hundreds of residues, or short structural element made of several residues. Also, there seems no signature sequence for domain swapping, since it is found that there is absence of sequence homology among those swapped proteins. As of now, it seems impossible to predict the propensity towards domain-swapping based upon sequence homology or secondary structure analysis (Dehouck, Biot et al. 2003). In addition, the interaction found in those swapped closed interface is also very diverse enough to include hydrophobic, hydrophilic, electrostatic or disulfide bond interaction. All the cases of domain-swapping tells us that each closed interface is stabilized by a variety of interaction as mentioned above and also seems to contribute to the energy required for the disruption of the closed interface during domain-swapping.

*Flexibility and diversity of the hinge loops*

When it comes to analyzing domain swapping, there must exist a hinge loop, which can function as beginning point of entire swapping and breaking point (Rousseau, Schymkowitz et al. 2003). Without the existence of such an intrinsic flexible hinge region, swapping event must be prevented. Based upon the previous domain-swapped structures, a hinge loop is shown to adopt various conformations in intact monomer and in the swapped oligomer. For example, the structural analysis of RNase swapped dimer strongly suggests that hinge loop region is intrinsically flexible, due to the presence of

variable hinge-loop conformations found in monomer and swapped-dimer (Jaskolski 2001).

Although we might assume that a hinge loop is flexible and unstructured loop, it turns out that it displays a variety of secondary structures in domina-swapped proteins. Some hinge loop regions are coils, some β-strands, and other s form α-helix. As mentioned earlier, the same hinge loop can take different conformations. In the RNase A, a hinge loop can form either a coil or a helix.

# Chapter 5: Conclusions and Future Directions


The crystallographic analysis reveals unexpected finding that OSR1 kinase domain is dimerized in domain-swapped manner, which is considered a novel mode of protein-protein interaction in the protein kinase family. Given the clear electron density map derived from linking region between two subunits as well structural alignment analysis, it is considered that domain swapping in the OSR1 kinase structure is bona fide, not the product of crystallization artifact.

The detailed structural analysis tells us that αEF helix and P+1 loop region located in the activation segment were completely swapped between two monomers. Most of this swapped region corresponds to *Ste20* "signature" sequence. It thus follows that although it is not still clear, there might be some correlation between domain-swapping event and substrate recognition mechanism. It is also revealed that DFG magnesium binding loop and N-terminal F-helix seem to function as anchor for hinge loop, based upon the structural alignment with other domain-swapped Ste20p kinases, which were reported after OSR1 structure determination. Also, it shoud be noted that those swapped Ste20p kinases also form domain-swapped dimer in a strikingly similar manner to OSR1.

In addition, the multiple sequence alignment analysis between domain-swapped kinases suggests that the conserved sequence (**R-x-x-F**) located in the c-terminal region of activation loop might play a certain role, which remains unclear thus far. The sequence is uniquely conserved only among the domain-swapped protein kinases.

Another interesting observation is from polar interaction between catalytic lysine and tryptophan from P+1 loop, because this intermolecular interaction in the swapped region is well conserved among the domain-swapped protein kinases from Ste20 family. This polar interaction between swapped polypeptide is characterized by cation-$\pi$ interaction, which is frequently found in other non-kiase domain-swapped protein structures.

However, as this interaction is also found in the same location from non-swapped Ste20 PAK1 or TAO2, it is yet to be investigated how this cation-$\pi$ interaction contributes to domain-swapping in the protein kinase.

Despite the detailed analysisi of OSR1 kinase structure, however, the underlying mechanism of domain swapping still remains elusive. Thus, a variety of questions should be answered on whether domain-swapping is coupled to reversible signaling event, what the consequene of domain swapping is, and how this domain swapping is triggered. Most of all, the determination of non-swapped OSR1 kinase structure will provide significant insight into its underlying story. Thus, by comparing the swapped species with non-swapped one from the same molecule, more complete picture of domain-swapping in the protein kinase can be drawn.

# Chapter 6: Expression, purification and crystallization of human Aurora B and INCENP

## Introduction

The success of mitosis depends to a large degree on the integration of chromosomal and cytoskeletal behavior. Much is now known about the components of the spindle assembly checkpoint, which monitors the attachment of chromosomes to the spindle and ensures that anaphase onset is delayed until the process is complete. However, less is known about the mechanism coordinating the behavior of the chromosomes, the mitotic spindle and the cleavage furrow during the closing stages of mitosis.

Chromosomal passengers are proteins that move from centromeres to the spindle midzone during mitosis (Earnshaw and Bernat 1990; Biggins 1999; Bishop and Schumacher 2002; Chen 2003). Recent studies show that the passenger inner centromere protein (INCENP) and Aurora-B kinase are in a complex that might also contain a third passengers, surviving (Burke and Stukenberg 2003; Carmena and Earnshaw 2003; Higuchi and Uhlmann 2003; Honda, Korner et al. 2003). The chromosomal passenger complex functions throughout mitosis in chromosome condensation and segregation, and at the end of mitosis, in the completion of cytokinesis (Earnshaw and Bernat 1990; Bolton 2002; Cheeseman 2002). As a component of the "chromosomal passenger protein complex," the Aurora B kinase is associated with centromeres during prometaphase and

105

with midzone microtubules during anaphase and is required for both mitosis and cytokinesis.

Aurora-B, as a component of chromosomal passengers, is related to a protein kinase, first identified in Drosophila Melanogaster in a search for genes that regulate the structure and function of mitotic spindle (Earnshaw and Bernat 1990; Bolton 2002; Cheeseman 2002). Ablation of Aurora B causes defects in both prometaphase chromosomal congression and the spindle checkpoint. However, the mechanisms underlying these defects are unclear. Also, Aurora kinases consisting of A, B, C have generated a lot of interest because of a link with cancer. The gene encoding Aurora-A is amplified in several human cancers, and Aurora-B and –C are overexpressed in overexpressed in many human cancer cell lines (Cheetham 2002). To investigate the functional and structural aspects of Aurora B kinase as a component of passenger protein complex relating to chromosomal segregation, the purification and crystallization of this protein has been started for X-ray crystallography as well as the relevant enzymatic and biophysical assays.

## Materials and Procedures

A series of constructs were designed and subcloned into pHis-parallel vector containing N-terminal 6X histidine-tag for both human Aurora B kinase and INCENP. Due to the ample evidence revealing that Aurora B kinase forms a stable complex with INCENP in intracellular environment, extensive trials for expressing and crystallizing their complex structure were also given, by utilizing coexpression vector such as petDuet

vector (Novagen) for *E.coli*. In order to produce soluble target proteins as monomer or complex, a series of constructs of varying fragment was generated for both Aurora B and INCENP.

Another important component of chromosomal passenger complex, survivin, was also expressed and purified. However, due to the lack of biophysical evidence showing that survivin is actually involved in this chromosomal passenger complex via stable protein-protein interaction, it was suggested that crystallographic studies of surviving-involving intermolecular contact in this complex might be risky and infeasible. It thus follows that our efforts were focused upon expression of separate Aurora B and INCENP, or Aurora B complexed with INCENP.

Although the expression trials for Auroa B and INCENP were given also in insect cells in collaboration with Hongtao Yu lab (UTSW), the amount of protein expressed was so low that additional effort for using Sf9 cells as expression system was discontinued.

**Cloning of Constructs**

Each construct was subcloned into the pHis-parallel vector containing N-terminal 6X histidine-tag. As mentioned earlier, petDuet vector for expressing complex structure was also utilized. All the constructs were generated based upon previous sequence-domain analysis. The PCR reaction condition I used is as follows.

2 μl     template

1 μl     primer 1 (50pmol/μl)

1 μl     primer 2 (50pmol/μl)

5 µl    *Pfu* reaction buffer (10X)

1 µl    *Pfu* enzyme

1 µl    dNTP (10mM) mixture

39 µl   water_____

50 µl   total reaction volume

This mixture was applied to PCR reaction to amplify the target sequence.

1st Cycle:              94 °C for 2 minutes to denature the template

2nd ~ 30th Cycle:    94 °C for 30 seconds to denature the template

                        58 °C for 40 seconds for primer annealing

                        72 °C for 90 seconds for primer extension

After PCR reaction, the PCR product was checked by DNA electrophoresis and purified. And this PCR product was digested by restriction enzymes as follows and applied to ligation reaction by T4 DNA ligase, followed by DNA sequencing.


30 µl PCR (purified)

2 µl restriction enzyme I

2 µl restriction enzyme II

0.5 µl BSA (100X)

5 µl restriction enzyme buffer

10.5 µl water_____

50 µl  Total reaction volume

This mixture was applied to digestion reaction at 37 °C for 2 hrs, followed by subsequent purification for ligation procedure.

**Expression of human Aurora B kinase and INCENP in *E.coli***

The basic expression procedures for individual Aurora B and INCENP are the same as one for Aurora B complexed with INCENP. Here, I described the standard procedure for expressing human Aurora B kinase.

A. Transformation

1. Mix 20 µl of Rosetta (DE3) cells or other *E.coli* strain with 1 µl of DNA plasmid vector and incubate the mixture on ice for 5 minutes

2. Heat-shock was applied to cell-DNA mixture for 40 sec by using 42°C water bath.

3. Transfer the cell-DNA sample onto ice promptly and keep it on ice for 3 minutes.

5. Add 80 µl of sterilized LB or SOC medium and grow the cells for 1 hour at 37°C in the temperature-controlled shaker

5. Apply the grown cells onto the LB plate containing

B. Cell Growth

1. Streak out *E. coli* strain Rosetta (DE3) transformed with the pHis-pararell vector encoding human Aurora B kinase on a LB plate containing 100 µg/ml of ampicillin (or 50 µg/ml of carbenicillin).

2. Incubate the LB plate overnight at 37°C to grow the transformed colonies.

3. Inoculate 100 ml of LB medium containing 100 µg/ml of ampicillin in a 250 ml flask with a single colony, and grow the cells at 37°C in a shaker

(250~300 rpm) overnight.

4. Use 10 ml of the overnight-grown culture to inoculate 1L of sterilzed LB medium containing 100 μg/ml of ampicillin.

5. Grow the cells at 37°C until OD at 595 nm reaches about 0.8.

6. Add 0.5 ~ 1 mM IPTG to induce kinase Aurora B kinase expression and grow the cells at 16 °C or 25 °C for overnight.

7. Harvest the cells by using a Beckman J6-MI centrifuge with a 6x1 L swing-out rotor for 30 min at 4000 rpm (at 4°C).

8. Resuspend the cell pellet with about 35 ml of chilled Ni-binding buffer and treat it with the protease inhibitor cocktail.

9. Flash-freeze the resuspended cell pellet in liquid nitrogen and store in – 80°C freezer.

**Purification of human Aurora B kinase and INCENP**

A. Preparation of the Cell-free Extract (from *E.coli*)

1. Thaw the frozen cells (taken from freezer-storage) at 4°C cold room or on ice.

2. Following the treatment of cell lysate with lysozyme and bezonase nucleases at 4°C for about 1 hr, sonicator was used to break the thawed cells on ice for about 3 minutes (10 sec on and 5 sec off)

3. By utilizing centrifugation of the cell lysate at 35000 rpm for 1 hr (4°C), the spinned-down cell debris was separated from the supernatant cell extract containing the target protein.

4.  The cell extract cleared by 0.45 μm syringe filter was transferred to the
    prechilled beaker for the next purification step.

B.  Immobilized Metal Affinity Chromatography

1.  By using a peristaltic pump (Amersham Biociences), the cleared cell
    extract was applied to a fast-chelating sepharose column (Amersham
    Pharmacia) loaded with 0.1 mM NiSO and pre-equilibrated with 5~10
    column volume of Ni-binding buffer.

2.  In order to remove non-specific bound proteins, after loading the sample
    (the cleared cell extract), the column was washed by 3~4 column volume
    of washing buffer with the increasing concentration of imidazole (from 10
    to 40 mM) using step gradient.

3.  Elute the target protein with 250 mM imidazole-containing Ni-elution
    buffer and pool the eluted fractions.

C.  TEV-protease Digestion

1.  The eluted sample fractions from Ni-column was promptly was mixed with
    the MonoQ A buffer and transferred into dialysis bag for buffer change at
    4°C for overnight.

2.  Additionally, about 1/30 ~ 1/50 ratio of TEV-protease was added for tev-
    protease digestion for removing N-terminal 6X His tag.

D.  Anion Exchange Chromatography

1. Spin down the dialyzed protein solution to remove the precipitated debris using 4000 rpm centrifugation.

2. Clear the dialyzed protein solution by 0.22 μm filter syringe

3. Load the filtered protein sample onto a MonoQ anion-exchange column (Pharmacia).

4. Following the washing step with the 3 column-volume of MonoQ A buffer, the protein sample was applied to step gradient for elution of target protein sample.

E. Size-Exclusion Chromatography

1. The eluted protein sample from anion-exchange chromatography was pooled and concentrated up to the total volume of 2~3 ml.

2. Following the filtration of sample protein by 0.22 μm filter syringe, the filtrate was loaded onto Superdex-75 or 200 size-exclusion column. The eluted sample was pooled and collected. The protein sample for crystallization was usually concentrated up to ~ 7 mg/ml.

**Crystallization of human Aurora B kinase and INCENP**

Every soluble construct produced was concentrated and screened for crystallization by utilizing commercially available sparse matrix kits (Emerald Biosciences and Hampton Research). The hanging-drop vapor diffusion method was used to grow crystals at 16 °C.

# Results and Discussion

**Expression of human Aurora B kinase and INCENP**

Although the expression level of human Aurora B kinase from Rosetta (DE3) was high, a variety of INCENP fragments subcloned into pHis-Parallel vector was not expressed. Based upon the assumption that the low expression level of INCENP fragment might be originated from the absence of stabilizing binding partner like Aurora B kinase, coexpression trials were also given extensively. However, any combinations of coexpression failed to elevate the expression of INCENP of any fragments. Thus, only Aurora B kinase domain was available for further purification.

**Purification of human Aurora B kinase**

Human Aurora B kinase domain was purified following the regular purification procedures from Ni column to gel filtration. It turned out that Aurora B kinase was not stable, also highly prone to precipitation. Right after Ni metal affinity chromatography, it was found that the eluted Aurora B kinase sample began to be precipitated heaviliy. A wide range of modifications for producing and separating this target protein was made extensively in terms of expression and purification, although it did not help. As mentioned earlier, the insect cell was used but its expression level was low and also unstable. Thus, no further improvement has been made with regard to its expression and solubility since then.

# Chapter 7: Expression, purification and crystallization of

## *Ste20 Hippo* kinase and Sav

## Introduction

The balance between cell death and cell proliferation is essential to coordinate homeostasis within multicellular organisms. The recently identified *hippo* encodes a Ste-20 family protein kinase that binds to and phosphorylates the tumor suppressor protein Salvador (Sav), which is known to interact with the Warts (Wts) protein kinase (Wu, Huang et al. 2003). Thus, *hippo* can be involved in regulating both cell proliferation and cell death in Drosophila.

*Hippo* encodes a polypeptide with an N-terminal kinase domain and a noncatalytic C-terminal domain. Its kinase domain belongs to Ste20 family Ser/Thr kinase. Ste20 family is a putative yeast mitogen-activated protein kinase kinase kinase kinase (MAP4K) involved in mating pathways (Dan, Watanabe et al. 2001). The Ste20 family kinases are further classified into the p21-activated kinases (PAK) and germinal center kinase (GCK) subfamilies.

The overall architecture and catalytic kinase domain sequence places *Hpo* into the group II GCK subfamily (Dan, Watanabe et al. 2001). Hpo is most closely related to human proteins MST2 (60% identity) and MST1 (58% identity) (Creasy and Chernoff 1995). MST1 and MST2 were first isolated based on their homology to yeast Ste20 and

independently identified as kinases activated in NIH3T3 cells by extreme cells (Taylor, Wang et al. 1996; Wu, Huang et al. 2003).

According to the current model, the C-terminal non-catalytic domain of *Hpo* associates with Sav and presents Sav to the *Hpo* kinase.  Also, it is suggested that Hpo associates with and phosphorylates Sav and interactions between *Hpo* and Sav facilitate Wts phsophorylation by *Hpo*.  This model is consistent with the previous observation that there exist direct physical interactions between Sav and Wts.  Thus, Sav can be viewed as an adaptor prtotein that brings *Hpo* in proximity to Wts to facilitate Wts phosphorylation.

Whereas the Sav WW domains have been implicated in Sav/Wts interaction, there has been speculation that the coiled-coil domain of Sav, located C-terminal to the WW domains, might be involved in Sav/*Hpo* interaction, as shown in Figure 7-1.  Thus, it is proposed that *Hpo*, Sav, and Wts define a tumor suppression pathway that coordinately regulates cell proliferation and apoptosis, and the Hpo-Sav-Wts pathway might be involved in tumorigenesis in mammals.

The coiled-coil region located c-terminal to WW domain is also found in Rassf proteins, as well as Hpo and Sav.  As shown in **Figure 7-2**, these coiled-coil regions from each protein reveal noticeable sequence homology characteristic of generic but slightly different coiled-coil heptad structure.  This new homology domain is termed as SARAH (Sav/Rassf/Hpo) domain (Scheel and Hofmann 2003).  SARAH domain sequence alignment in **Figure 7-2** shows that there are no invariant residues but there are several highly conserved positions.  As indicated in **Figure 7-3**, although some of conserved positions can be explained by generic coiled-coil properties, some novel heptad arrangements are noticed from 'a'and'd' polar residues, which normally favor

hydrophobic side chanins. The presence of polar side chains in 'a' and 'd' positions is reminiscent of the SNARE tetramerization (Scheel and Hofmann 2003), which is characterized by its specific and reversible coiled-coil interaction. Also, other polar side chains found in 'g' and 'c' positions, adjacent to the predicted inter-helical contact surface, imply the possibility that these residues can be utilized for formation of higher-order coiled-coil interaction. Regulatory interactions mediated by coiled-coils usually involve relatively short helices capable of interacting reversibly in a homotypic and heterotypic manner. As indicated in **Figure 7-1**, SARAH domain mediates a heterotypic interaction relaying the signals from *Hpo* vai scaffold protein Sav to the downstream Wts. Also, SARAH domain can mediate homotypic interactions, as shown in the case of human *Hpo*-ortholog Mst1 dimerization by its c-terminal region. In addition, as the SARAH domain has coiled-coil characteristics with an unusual heptad arrangement, it is thought that SARAH is in fact a trimerization domain (Scheel and Hofmann 2003). It thus can be proposed that the SARAH domain plays a crucial role as reversible mediator characteristic of unusual heptad structure, unifying and merging both Ras and *Hpo* pathways involved in delicate coordination of cell death and proliferation.

**Figure 7-1. Putative Model of a unified Hippo (Mst) kinase-based pro apoptotic pathway.**
(A) The Hpo Sav Wts pathway (B) The Ras Rassf Mst1 pathway (C) A unified model based upon the proposed SARAH trimerization. The letter 'P' indicates the PPxY motifs recognized by WW domains. This diagram was modified from (Scheel and Hofmann 2003).

```
Hs WW45        321 ILKWELFQLADLDTYQGMLKLLFMKELEQIVKMYEAYRQALLTELENR
Dm Salvador   552 LLQFNMFSLPELEGFDSMLVRLFKQELGTIVGFYERYRRALILEKNRR
Ce T10H10.3   282 LLEWDLFNFEQLTEYEHLMMKLYKQEVFDIVKKYEKKRNVLNREIHRR


Hs Rassf1     220 EVNWDAFSMPELHNFLRILQREEBEHLRQILQKYSYCRQKIQEALHAC
Hs Rassf2     272 VAQYIKFEMPVLKSFIQKLQEEEDREVKKLMRKYTVLRLMIRQRLEEI
Hs Rassf3     187 IGEWEAFSLPELQNFLRILDKEEDEQLQNLKRRYTAYRQKLEEALREV
Hs Rassf4     270 VAQYIKFEMPVLDSFVEKLKEEEREIIKLTMKFQALRLTMLQRLEQL
Hs Nore1      366 EVEWDAFSIPELQNFLTILEKEEQDKIQQVQKKYDKFRQKLEEALRES
Hs Rassf6     281 VAQYINFHFSLLESILQRLNEEEKREIQRIVTKFNKEKAIILKCLQNK
Dm CG46456    755 VAQFLNLSLPECRAILERYDQELAREVAKIKERYAELRRRIVSRMESL
Ce T24F1.3    498 DILWDAFEIPELENFLRILGMEEKQYVFQTQQKYQQYRYHLDAELRQR


Hs Mst1       433 YEFLKSWTVEDLQKRLLALDPMMEQEIEEIRQKYQSKRQPILDAIEAK
Hs Mst2       437 FDFLKNLSLEELQMRLKALDPMMEREIEELRQRYTAKRQPILDAMDAK
Dm CG11228    608 FEFLKFLTFDDLNQRLCNIDHEMELEIEQLNKKYNAKRQPIVDAMNAK
Ce F14H12.4   474 FEFLRNITLDELIRRKESLDSEMBEEIRELQRRYKTKRQPILDVIEIK
```

**Figure 7-2. Alignment of representative SARAH domains**

For all domain sequences, sequences from human (Hs), *Drosophila* (Dm) and *Caenorhabditis elegans* (Ce) are shown. This sequence alignment was regenerated and modified based upon (Scheel and Hofmann 2003).
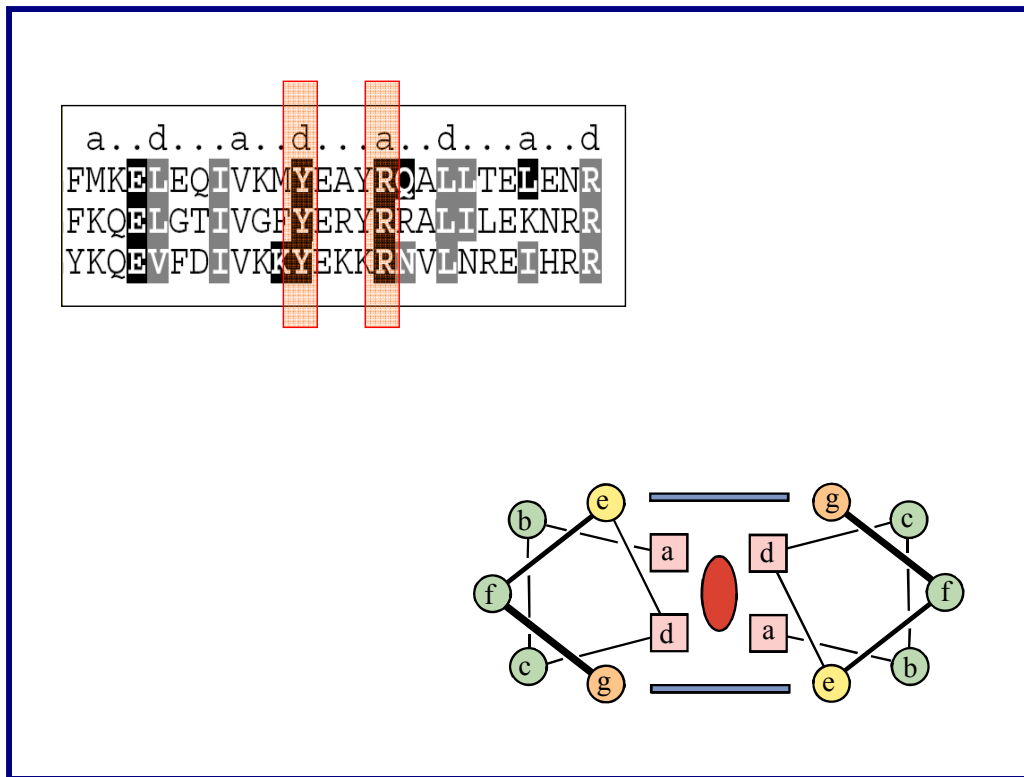
**Figure 7-3. SARAH domains : Unusual Heptad Repeat**

# Materials and Procedures

A series of constructs were designed and subcloned into pHis-parallel vector containing N-terminal 6X histidine-tag for both *Ste20p Hippo* and Salvador (Sav) as indicated in **Figure 7-4**. Especially, due to the size of *Ste20p Hippo*, various fragments of *Hippo* were subcloned and tested for its expression. In parallel, a variety of fragments from hippo and sav was subcloned into petDuet vector (Novagen) for their coexpression in *E.coli*.

Since there had been previous studies indicating that *Ste20p Hippo* interacts with Sav and forms a complex structure via SARAH domain, extensive trials for expressing and purifying hippo c-terminal domain with Sav fragment were also given. Although the expression trials for *hippo* and sav were given also in insect cells, the amount of protein expressed was so low that additional effort for using Sf9 cells as expression system was discontinued.

**Cloning of Constructs**

Each construct was subcloned into the pHis-parallel vector containing N-terminal 6X histidine-tag to express either Hippo or sav. petDuet vector (Novagen) was also utilized for simultaneous coexpression of two target proteins. All the constructs were generated based upon previous sequence-domain analysis. The PCR reaction condition I used is as follows.

2 μl     template

1 μl     primer 1 (50pmol/μl)

1 μl     primer 2 (50pmol/μl)

5 μl     *Pfu* reaction buffer (10X)

1 μl     *Pfu* enzyme

1 μl     dNTP (10mM) mixture

39 μl   water_____

50 μl   total reaction volume

This mixture was applied to PCR reaction to amplify the target sequence.

 1$^{st}$ Cycle:          94 °C for 2 minutes to denature the template

 2$^{nd}$ ~ 30$^{th}$ Cycle:    94 °C for 30 seconds to denature the template

                        58 °C for 40 seconds for primer annealing

                        72 °C for 90 seconds for primer extension

After PCR reaction, the PCR product was checked by DNA electrophoresis and purified.

And this PCR product was digested by restriction enzymes as follows and applied to

ligation reaction by T4 DNA ligase, followed by DNA sequencing.

30 μl PCR (purified)

2 μl restriction enzyme I

2 μl restriction enzyme II

0.5 μl BSA (100X)

5 μl restriction enzyme buffer

10.5 μl water_____

50 μl  Total reaction volume

This mixture was applied to digestion reaction at 37 °C for 2 hrs, followed by subsequent

purification for ligation procedure.



**Figure 7-4.  Expression trials for Hippo and Sav in *E.coli*.**

**Expression of *Ste20p Hippo* and its binding partner Sav**

A.  Transformation

　　1.  Mix 20 µl of Rosetta (DE3) cells or other *E.coli* strain with 1 µl of DNA

　　　　plasmid vector and incubate the mixture on ice for 5 minutes

　　2.  Heat-shock was applied to cell-DNA mixture for 40 sec by using 42°C

　　　　water bath.

3. Transfer the cell-DNA sample onto ice promptly and keep it on ice for 3
   minutes.

6. Add 80 μl of sterilized LB or SOC medium and grow the cells for 1 hour at 37°C
   in the temperature-controlled shaker

5. Apply the grown cells onto the LB plate containing


B. Cell Growth

1. Streak out *E. coli* strain Rosetta (DE3) transformed with the pHis-parallel
   vector encoding *Ste20p hippo* on a LB plate containing 100
   μg/ml of ampicillin (or 50 μg/ml of carbenicillin).

2. Incubate the LB plate overnight at 37°C to grow the transformed colonies.

3. Inoculate 100 ml of LB medium containing 100 μg/ml of ampicillin in a
   250 ml flask with a single colony, and grow the cells at 37°C in a shaker
   (250~300 rpm) overnight.

4. Use 10 ml of the overnight-grown culture to inoculate 1L of sterilzed LB
   medium containing 100 μg/ml of ampicillin.

5. Grow the cells at 37°C until OD at 595 nm reaches about 0.8.

6. Add 0.5 ~ 1 mM IPTG to induce *Ste20p hippo* expression and grow
   the cells at 16 °C or 25 °C for overnight.

7. Harvest the cells by using a Beckman J6-MI centrifuge with a 6x1 L
   swing-out rotor for 30 min at 4000 rpm (at 4°C).

8. Resuspend the cell pellet with about 35 ml of chilled Ni-binding buffer and
   treat it with the protease inhibitor cocktail.

9. Flash-freeze the resuspended cell pellet in liquid nitrogen and store in –80°C freezer.

**Purification of *Ste20p Hippo* and Sav**

A. Preparation of the Cell-free Extract (from *E.coli*)

1. Thaw the frozen cells (taken from freezer-storage) at 4°C cold room or on ice.

2. Following the treatment of cell lysate with lysozyme and bezonase nucleases at 4°C for about 1 hr, sonicator was used to break the thawed cells on ice for about 3 minutes (10 sec on and 5 sec off)

3. By utilizing centrifugation of the cell lysate at 35000 rpm for 1 hr (4°C), the spinned-down cell debris was separated from the supernatant cell extract containing the target protein.

4. The cell extract cleared by 0.45 μm syringe filter was transferred to the prechilled beaker for the next purification step.

B. Immobilized Metal Affinity Chromatography

1. By using a peristaltic pump (Amersham Biociences), the cleared cell extract was applied to a fast-chelating sepharose column (Amersham Pharmacia) loaded with 0.1 mM NiSO and pre-equilibrated with 5~10 column volume of Ni-binding buffer.

2. In order to remove non-specific bound proteins, after loading the sample (the cleared cell extract), the column was washed by 3~4 column volume

of washing buffer with the increasing concentration of imidazole (from 10

to 40 mM) using step gradient.

3.  Elute the target protein with 250 mM imidazole-containing Ni-elution

    buffer and pool the eluted fractions.

C.  TEV-protease Digestion

1.  The eluted sample fractions from Ni-column was promptly was mixed with

    the MonoQ A buffer and transferred into dialysis bag for buffer change at

    4°C for overnight.

2.  Additionally, about 1/30 ~ 1/50 ratio of TEV-protease was added for tev-

    protease digestion for removing N-terminal 6X His tag.

D.  Anion Exchange Chromatography

1.  Spin down the dialyzed protein solution to remove the precipitated debris

    using 4000 rpm centrifugation.

2.  Clear the dialyzed protein solution by 0.22 µm filter syringe

3.  Load the filtered protein sample onto a MonoQ anion-exchange column

    (Pharmacia).

4.  Following the washing step with the 3 column-volume of MonoQ A buffer,

    the protein sample was applied to step gradient for elution of target protein

    sample.

E.  Size-Exclusion Chromatography

1.  The eluted protein sample from anion-exchange chromatography was

pooled and concentrated up to the total volume of 2~3 ml.

2. Following the filtration of sample protein by 0.22 μm filter syringe, the filtrate was loaded onto Superdex-75 or 200 size-exclusion column. The eluted sample was pooled and collected. The protein sample for crystallization was usually concentrated up to ~ 12 mg/ml.

**Crystallization of *Ste20p Hippo* and Sav**

Every soluble construct produced was concentrated and screened for crystallization by utilizing commercially available sparse matrix kits (Emerald Biosciences and Hampton Research). The hanging-drop vapor diffusion method was used to grow crystals at 16 °C. For protein crystallization trials, a variety of concentration was tested from 10 mg/ml to 2 mg/ml.

# Results and Discussions

**Expression of *Ste20p Hippo* and Sav**

Generally, the expression level of *Ste20p Hippo* fragments containing kinase domain was very low in *E.coli*. However, non-kinase domain of *Ste20p Hippo* (of varying lengths), c-terminal region, was expressed in high level from *E.coli* strain using Rosetta (DE3) cells. The expression of Sav (of varying lengths) was virtually zero based upon SDS-PAGE analysis. Among all the constructs I generated, *Ste20p Hippo* C-terminal (474~669) fragment gave highest level of protein expression and stability.

**Purification of *Ste20p Hippo* (474~669)**

As explained above briefly, hippo c-terminal (474~669) fragment was the only protein sample available for subsequent purification and crystallization, in terms of its yield and stability. By using standard protein purification steps, Ni affinity, anion-exchange, size-exclusion chromatography were applied in order. Throughout the entire purification step, there was no sign of aggregation. Interestingly enough, but not surprisingly, this construct (474~669) containing the SARAH domain, which is known as oligomerization motif for Hippo-based pro-apoptotic pathway, made a tetramer formation during size-exclusion purification step. The peak from gel filtration corresponds to ~ 80 kDa size of protein sample but the actual molecular weight of this construct (474~669) is about 20 kDa as indicated in Figure 7-5 and Figure 7-6. The final protein sample was highly pure based upon SDS-PAGE analysis.
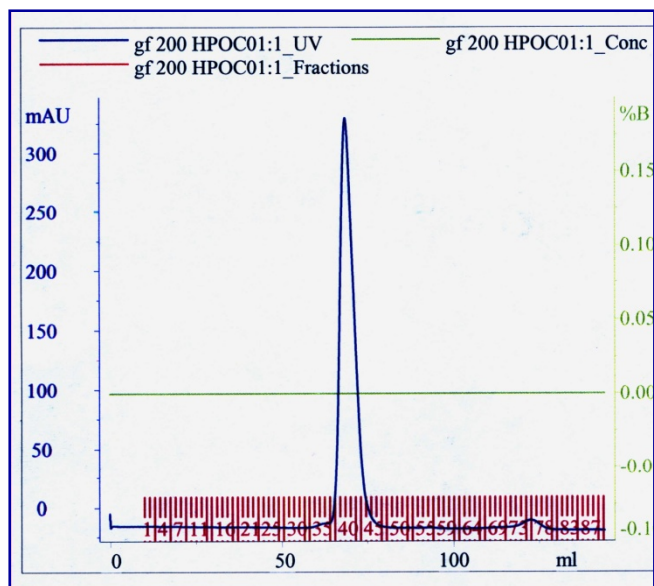
**Figure 7-5. Purification of Hippo (474-669) by size exclusion chromatography** This c-terminal Hippo construct (474-669) forms a tetramer during size-exclusion chromtagraphy. Whereas this single peak corresponds to ~80 kDa size of protein, the actual molecular weight of this hippo constructs is about ~20 kDa.
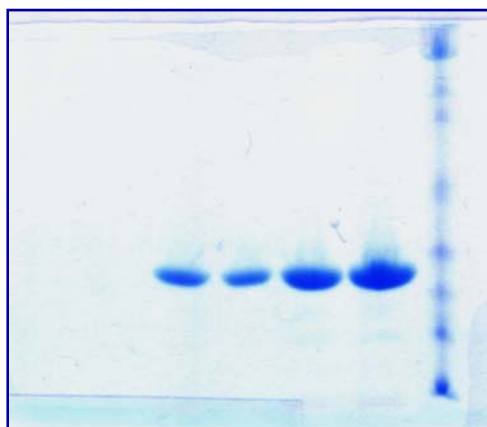


**Figure7-6. SDS-PAGE Analysis of Hippo (474-669) after size-exclusion chromatography.** This SDS-PAGE analysis indicates the hippo construct finally purified from gel filtration is highly pure. These four samples were just repeated after size-exclusion chromatography.

**Crystallization of *Ste20p Hippo* (474~669)**

Since *Ste20p Hippo* c-terminal construct (474~669) was only protein sample available for crystallization, extensive efforts were focused on the establishment of initial condition. As always, commercially available crystallization screening kits were utilized. However, initial trials gave no clues in terms of optimized crystallization condition, even after I tried various protein concentration and also different buffer pH level.

Finally, temperature-controlled crystallization was tried for general screening. This temperature-controlled experiment involved heating of protein sample at 50 °C in water bath for 10 minutes, prior to its crystallization set-up. Surprisingly, this heat-treatment made some noticeable improvement in terms of crystallization behavior. This improvement in crystallization derived from heat-utilizing reannealing method involved the intriguing transition from quasi-crystal form to spherulitic forms of crystals over the various conditions from commercialliy available screening kits as clearly revealed in **Figure 7-7** and **Figure 7-8**. A variety of time period for heating was checked for this effect and several different temperatures from 30 °C to 50 °C were also tried. This transition from quasi- to spherulitic crystallization was found to necessitate the temperature over 40 °C and heating time more than 5 minutes. Interestingly, however, this improvement in crystallization behavior was only observed at low protein concentration (about 2 mg/ml). But no further improvement was made in terms of protein crystallization.
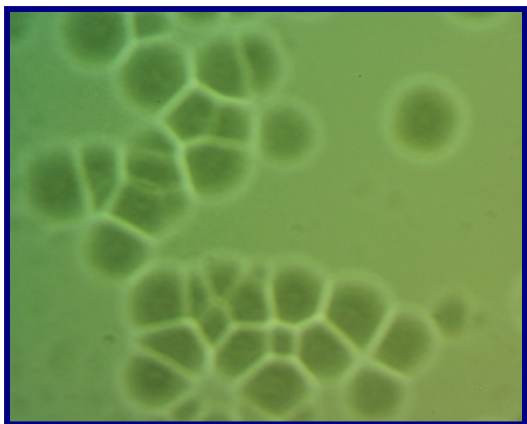
**Figure 7-7.  Initial quasi crystal formation from *Ste20p Hippo* (474~669)**
A set of crystallization condition generated this kind of quasi-crystals at low or high protein concentration (~2 mg/ml or 10 mg/ml).  The protein samples used for this crystal set-up were not treated with heating.
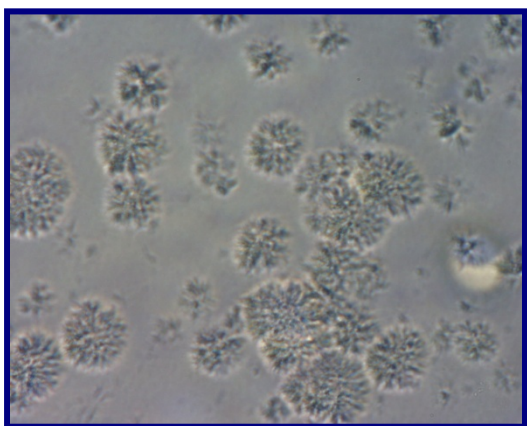


**Figure 7-8.  Improved crystallization of Ste20p Hippo (474~669) from**
**Heating.**   The protein samples used for this crystallization trial were heated for 10 minutes at 50 °C, giving rise to the remarkable improvement in crystal quality.

# Chapter 8: Expression, purification and crystallization of human ISG15

## Introduction

Ubiquitin is a well known protein modifier that changes the other protein's function and its final destination. Conjugation of ubiquitin to target protein is known to regulate a series of cellular processes including intracellular protein degradation, cell cycle, signal transduction, transcription, membrane trafficking and antigen presentation (Hershko and Ciechanover 1998; Hicke 2001).

A group of small proteins, called ubiquitin-like molecules (Ubls), are structurally related to ubiquitin and shown to be conjugated to different target proteins in a similar fashion as ubiquitin (Jentsch and Pyrowolakis 2000). Thus far, several Ubls like SUMO, NEDD8, Fub, Fat10, Apg12, and ISG15, have been characterized. Among those Ubls, ISG15, which was identified more than 20 years ago, is expressed highly in most cell types when infected by viruses or bacteria. The exact biochemical functions, however, are still unknown so far.

### ISG15, a member of ubiquitin-like proteins

ISG 15 was originally identified as an interferon stimulated gene (ISG), whose expression is stimulated upon interferon treatment (Farrell, Broeze et al. 1979; Korant, Blomstrom et al. 1984) as shown in **Figure 8-1**. This protein is also called ubiquitin

131

cross-reactive protein (UCRP) due to its cross-reactivity with ubiquitin antibodies (Loeb and Haas 1992). ISG15 consists of two ubiquitin homology domain in tendem (Farrell, Broeze et al. 1979) and shows 33% (N-terminal domain) and 32% (C-terminal domain) identity to ubiquitin in mice. Like other Ubls, ISG15 is translated as 17-kDa precursor and processed to 15-kDa functional unit by a protease cleavage reaction, which makes diglycine residues exposed at the c-terminus critical for subsequent conjugation to target proteins.

**Enzymes in ISGylation and deISGylation of cellular proteins**

Generally, it is expected that the mechanism for ISGylation and deISGylation is similar to that of ubiquitin. Thus far, the known components for ISG system include UBE1L (ISG15-activating enzyme) and UBP43 (deISGylating enzyme) (Kok, Hofstra et al. 1993). It is known that these components share a significant homology in their primary structure with their counterparts of ubiquitin system. Based upon its homology to ubiquitin-specific E2s, a set of interferon inducible proteins were suggested as an E2 for ISGylation (Pru, Austin et al. 2001). UBE1L was cloned as a gene encoding an enzyme homologous to the ubiquitin-activating enzyme E1 (Kok, Hofstra et al. 1993). A recent study shows that UBE1L is an E1 for protein ISG15 modification. UBP43 was originally cloned during the analysis of genes differentially expressed between wild type and leukemia fusion protein AML1-ETO knock-in mice (Yergeau, Hetherington et al. 1997). The primary amino-acid sequence indicates that UBP43 is a member of the UBP family of ubiquitin-specific proteases. UBP43 consists of the conserved domains, such as the Cys and His domains that are present in all UBP family members and expected to be

an ubiquitin-specific protease (Yergeau, Hetherington et al. 1997). UBP43 has enzymatic specificity towards ISG15. It also cleaves ISG15-peptide fusion instead of ubiquitin fusion and also removes ISG15 from native cojugates.

**ISG15 as a cytokine**

It is known that ISG15 is detected in cells both in the free and conjugated forms, upon interferon treatment (Loeb and Haas 1992). It is also detected as an extracellular (secreted or released) free protein (Knight and Cordova 1991). Secretion of free ISG15 from human monocytes and lymphocytes was also reported (Knight and Cordova 1991). In those human monocytes and lymphocytes, free ISG15 can function as a cytokine and stimulate interferon-γ production by CD3[+], but not by CD14[+] or CD56[+] cells (Recht, Borden et al. 1991). In the presence of CD3[+] cells, ISG15 induce the proliferation of CD56[+] natural killer cell and dramatically enhances non-major histocompatibility complex-restricted cytotoxicity (Knight and Cordova 1991).
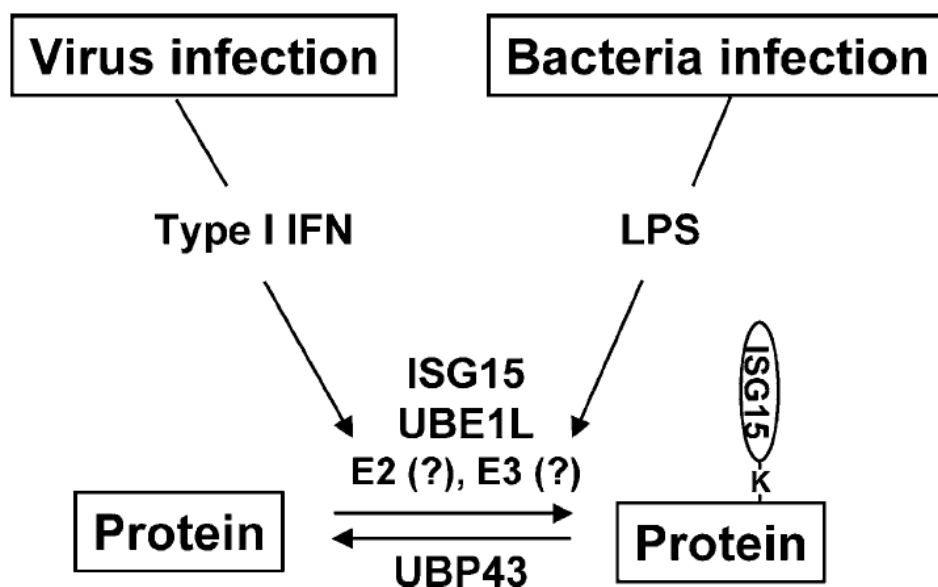
**Figure 8-1. The protein ISGylation system.**
Unlike other Ubl systems, protein ISGylation is stimulated by viral or bacterical infection. Upon viral infection, expression level of ISG15, UBE1L and UBP43 is highly elevated through type I interferon (IFN) signaling, giving rise to modification of cellular protein by ISG15. Bacterial infection, especially gram-negative bacteria, also activates the ISGylation system via the TLR4 signaling pathway which recognizes the bacterial cell surface molecule, lipopolysaccharide (LPS). The Figure 8-1 was adapted from (Kim and Zhang 2003).

# Materials and Procedures

As explained earlier in Introduction, there are two forms of ISG15, which include inactive uncleaved 17-kDa ISG15 and active cleaved 15-kDa ISG15. For crystallization trials, two forms of constructs were generated and expressed in high level from *E.coli*. These constructs were generated and provided by Gabriel Pineda (James Chen Lab, UTSW). The DNA fragments, which encode different forms of human ISG15, were

subcloned into pet14b vector with N-terminal 6X Histidine tag followed by thrombin cleavage site. Both constructs were expressed and purified for crystallization. I followed standard protein expression procedure for *E.coli*.

**Expression of human ISG15**

A. Transformation

 1. Mix 20 µl of BL21 (DE3) pLys cells with 1 µl of DNA plasmid vector and incubate the mixture on ice for 5 minutes

 2. Heat-shock was applied to cell-DNA mixture for 40 sec by using 42°C water bath.

 3. Transfer the cell-DNA sample onto ice promptly and keep it on ice for 3 minutes.

 7. Add 80 µl of sterilized LB or SOC medium and grow the cells for 1 hour at 37°C in the temperature-controlled shaker

 5. Apply the grown cells onto the LB plate containing ampicillin

B. Cell Growth

 1. Streak out *E. coli* strain BL21 (DE3) pLys transformed with the pet14b vector encoding ISG15 on a LB plate containing 100 µg/ml of ampicillin (or 50 µg/ml of carbenicillin).

 2. Incubate the LB plate overnight at 37°C to grow the transformed colonies.

 3. Inoculate 100 ml of LB medium containing 100 µg/ml of ampicillin in a 250 ml flask with a single colony, and grow the cells at 37°C in a shaker (250~300 rpm) overnight.

4. Use 10 ml of the overnight-grown culture to inoculate 1L of sterilzed LB medium containing 100 μg/ml of ampicillin.

5. Grow the cells at 37°C until OD at 595 nm reaches about 0.8.

6. Add 0.5 mM IPTG to induce ISG15 expression and grow the cells at 25 °C for overnight

7. Harvest the cells by using a Beckman J6-MI centrifuge with a 6x1 L swing-out rotor for 30 min at 4000 rpm (at 4°C).

8. Resuspend the cell pellet with about 35 ml of chilled Ni-binding buffer and mix it with the protease inhibitor cocktail.

9. Flash-freeze the resuspended cell pellet in liquid nitrogen and store in – 80°C freezer.


**Purification of human ISG15**

A. Preparation of the Cell-free Extract

1. Thaw the frozen cells (taken from freezer-storage) at 4°C cold room or on ice.

2. Following the treatment of cell lysate with lysozyme and bezonase nucleases at 4°C for about 1 hr, sonicator was used to break the thawed cells on ice for about 3 minutes (10 sec on and 5 sec off)

3. By utilizing centrifugation of the cell lysate at 35000 rpm for 1 hr (4°C), the spinned-down cell debris was separated from the supernatant cell extract containing the target protein.

4. The cell extract cleared by 0.45 μm syringe filter was transferred to the

prechilled beaker for the next purification step.

B. Immobilized Metal Affinity Chromatography

1. By using a peristaltic pump (Amersham Biociences), the cleared cell extract was applied to a fast-chelating sepharose column (Amersham Pharmacia) loaded with 0.1 mM NiSO and pre-equilibrated with 5~10 column volume of Ni-binding buffer.

2. In order to remove non-specific bound proteins, after loading the sample (the cleared cell extract), the column was washed by 3~4 column volume of washing buffer with the increasing concentration of imidazole (from 10 to 40 mM) using step gradient.

3. Elute the target protein with 250 mM imidazole-containing Ni-elution buffer and pool the eluted fractions.

C. Thrombin-protease Digestion

1. The eluted sample fractions from Ni-column was promptly was mixed with the MonoQ A buffer and transferred into dialysis bag for buffer change at 4°C for overnight.

2. Additionally, about 1/30 ~ 1/50 ratio of thrombin-protease was added for removing N-terminal 6X His tag.

D. Anion Exchange Chromatography

1. Spin down the dialyzed protein solution to remove the precipitated debris

using 4000 rpm centrifugation.

2. Clear the dialyzed protein solution by 0.22 μm filter syringe

3. Load the filtered protein sample onto a MonoQ anion-exchange column (Pharmacia).

4. Following the washing step with the 3 column-volume of MonoQ A buffer, the protein sample was applied to step gradient for elution of target protein sample.

E. Size-Exclusion Chromatography

1. The eluted protein sample from anion-exchange chromatography was pooled and concentrated up to the total volume of 2~3 ml.

2. Following the filtration of sample protein by 0.22 μm filter syringe, the filtrate was loaded onto Superdex-75 or 200 size-exclusion column. The eluted sample was pooled and collected. The protein sample for crystallization was usually concentrated up to ~ 30 mg/ml or higher.

**Crystallization of human ISG15**

Every soluble construct produced was concentrated and screened for crystallization by utilizing commercially available sparse matrix kits (Emerald Biosciences and Hampton Research). The hanging-drop vapor diffusion method was used to grow crystals at 16 °C.

# Results and Discussions

**Expression of human ISG15**

The expression level of human ISG15 (cleaved and uncleaved forms) in *E.coli* was high enough for subsequent protein purification. The same expression procedure using *E.coli* strain BL21 (DE3) pLys was applied to the two forms of constructs.

**Purification of human ISG15**

The two forms of human ISG15 (cleaved and uncleaved) were purified by the same purification method. After Ni metal affinity chromatography is applied, thrombin-cleavage digestion for removing histidine-tag was made during buffer change, followed by anionic-exchange and size-exclusion chromatography. The final buffer condition for purified ISG15 contains Hepes 20mM pH=7.5, NaCl 100 mM and DTT 2mM. Interestingly, it was found later from gel filtration purification step that some population of ISG15 is dimerized in the presence of reducing agent like DTT (2 mM). However, even in the presence of reducing agent, the majority of ISG15 from size-exclusion chromatography was still monomer. Although those two protein constructs were soluble and stable enough for final gel filtration, each purification step gave certain level of precipitation, requiring additional filitering procedure for removing aggregated species. Especially, during thrombin-cleavage digestion, precipitation level was highest. As shown in Figure 8-2, the final purified ISG15 had additional impure bands, which were not removed eventually.
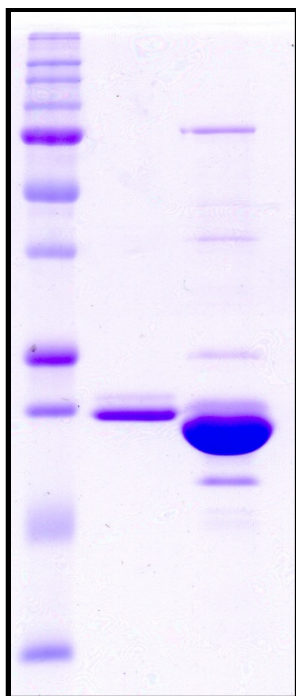
**Figure 8-2. SDS-PAGE Analysis of ISG15 purification**
The left lane and the right lane indicate uncleaved and cleaved form of ISG15, respectively, after final size-exclusion chromatography.

**Crystallization of human ISG15**

By utilizing commercially available protein crystallization screening kits, extensive efforts were focused upon finding our initial conditions for cleaved or uncleaved form of human ISG15. However, only cleaved form of ISG15 gave some crystallization conditions as shown in Figure 8-3 and Figure 8-4. Based upon these initial crystallization conditions, extensive efforts were made to improve quality and size of these ISG15 crystals as shown in Figure 8-5~8. However, the size of these improved crystals was too small (less than ~ 50 μm in the longest dimension) and they didn't diffract well.

**Figure 8-3. Initial crystallization of human ISG15.**
The crystals were grown by hanging drop vapor diffusion method. The crystals were grown at 16°C for 3 days. This initial crystallization condition contains 0.1M HEPES pH=7.5, 20% v/v Jeffamine M- 600 reagent
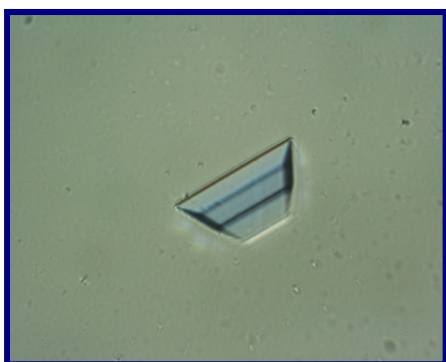


**Figure 8-4**. **Initial crystallization of human ISG15.**
The crystals were grown by hanging drop vapor diffusion method. The crystals were grown at 16°C for 3 days. This initial crystallization condition contains 1.0 M $(NH_4)_2HPO_4$ imidazole pH 8.0 100 mM.

**Figure 8-5**. **Improved crystallization of human ISG15**
This improved crystallization condition contains 1.0 M (NH4)2HPO4 imidazole pH 8.0 100 mM, in the presence of 50% v/v PEG 400 as additive (~1/20 ratio to whole drop volume).

.
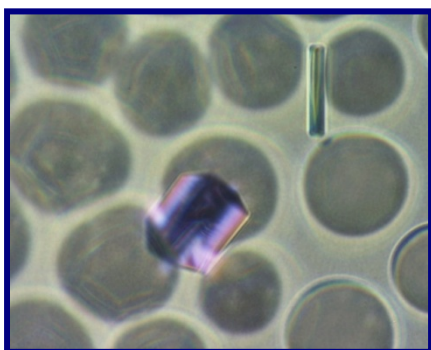


**Figure 8-6**. **Improved crystallization of human ISG15**
This improved crystallization condition contains 1.0 M (NH4)2HPO4 imidazole pH 8.0 100 mM, in the presence of 2.0 M NaCl as additive (~1/20 ratio to whole drop volume).

.
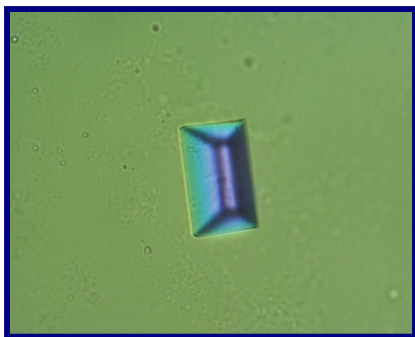
**Figure 8-7**.  **Improved crystallization of human ISG15.**
This improved crystallization condition contains 1.0 M (NH4)2HPO4 imidazole pH 8.0
100 mM, in the presence of 30% w/v Dextran Sulfate sodium salt as additive (~1/20 ratio
to whole drop volume).
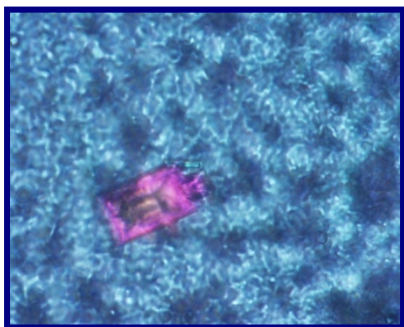


**Figure 8-8.  Improved crystallization of human ISG15.**
This improved crystallization condition contains 1.0 M (NH4)2HPO4 imidazole pH 8.0
100 mM, in the presence of 40% v/v 1.4 Butanediol as additive (~1/20 ratio to whole
drop volume).

# APPENDIX A

METHODS FOR ANALYZING MAPK CASCADES

## Crystallization of MAP kinase

SEUNG-JAE LEE[A], TIANJUN ZHOU[1, A] AND ELIZABETH J. GOLDSMITH

[a]Department of Biochemistry, The University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, TX 75390-8816, USA

## Abstract

X-ray structural studies of MAP kinases and MAP kinase module components are elucidating how kinase activity is regulated and how specificity of signaling is conferred. In the past decade, MAP kinases have been crystallized in their active, phosphorylated forms or low-activity, unphosphorylated forms, as well as in the presence of binding partners such as docking peptides and inhibitors. Crystallization has been achieved via diverse strategies including control of phosphorylation, coding sequence modification, incorporation of tags for purification, and use of a variety of cell-types for protein expression. Recently, interest has been focused on use of crystallography for lead optimization in the development for pharmacological inhibitors on MAP kinases. Further, some success has been gained in crystallizing the MAP kinase activators MAP2Ks and MAP3K kinase domains. This review describes the key methods that have been utilized to crystallize MAP kinases and MAP kinase pathway components.

**Article Outline**

## 1. Introduction

MAP kinase modules are three-tiered kinase cascades leading to the activation of MAP kinases. The modules are ubiquitous transducers of signals originating at the plasma membrane of eukaryotic cells leading to a range of outcomes including transformation, differentiation, and apoptosis. MAP kinase module components have become objects of extensive structural analysis for a variety of reasons. First is to understand how MAP kinases are activated by dual phosphorylation, and how MAP kinase modules carry out two double phosphorylation events [1] and [2]. Second, there are over twelve mammalian MAP kinase subfamilies, and each of the three well-studied MAP kinases ERK1/2, p38, and JNK are known to have several isoforms, each of which is also expressed in multiple splice-variants [3]. The mechanisms by which specificity is achieved among these enzymes are actively being studied through biochemical [4], [5] and [6] and structural analysis [7]. Third, MAP kinases and their activators are being pursued as drug targets for therapeutic intervention in a variety of diseases [8], [9] and [10]. The MAP kinase p38 has drawn the most extensive attention since its early identification as an anti-inflammatory agent [11], and owing to its role as a signal transducer responding to cellular stress stimuli such as cytokines. The recent studies show that MAP kinase p38 plays a role in initiating and sustaining the inflammatory response in

the brain [12], and thus may prove to be a drug target in the treatment of neurodegenerative diseases. ERK1/2 are major targets of Ras-dependent signals and are usually most strongly activated by growth factors and proliferative stimuli. Their activating enzymes MEK1/2 are also attractive targets due to their restricted substrate specificity for ERK1/2 and are found highly active in a variety of primary human tumors including cancers of breast, lung, colon, pancreas, prostate and others [13]. The JNK pathway is activated by certain cytokines, mitogens, osmotic stresses and its activation is required for some forms of stress-induced apoptosis. The JNK pathway activation may contribute to the neuronal death and atrophy relating to Alzheimer's Parkinson's, Huntington's diseases, and stroke [14]. The RAF family of MAP3Ks has also drawn attention as a drug targets. Active BRAF mutants are found in most melanomas and in colorectal [15] and [16], ovarian [17], and thyroid [18] cancers.

Due to the significance of MAP kinases and its pathway module components as signaling molecules, their unique regulation and their potential as drug targets, they have been extensively investigated by crystallographic methods to further understand their structural basis of regulation and specificities. The crystal structures of low activity ERK2 [19], low activity p38α [20] and [21], and low activity JNK3 [22] have been reported (Table 1). In addition, active structures of ERK2 [23], p38γ [24], and MAP3K Tao2 [25] have also been reported.

Interestingly in the recent years, several crystal structures of MAP kinases complexed with docking peptides have been reported to help elucidate the structural basis of docking interaction in the MAP kinase pathway (Table 2). They include p38α bound with peptide derived from its activator and substrate [26], JNK1α1 complexed with peptide from JNK1-interacting protein 1 [27], Fus3 with several peptides from activating and inactivating enzymes as well as its substrate [28] and [29], and ERK2 complexed with peptide derived from MAP kinase phosphatase [30]. Also, numerous crystal structures of MAP kinases such as p38, ERK2, and JNK3 have been reported in complex with a diverse array of inhibitors (Table 3). The inhibitor-bound forms of crystal structures of human MEK1 and MEK2 as well as B-RAF (MAP3K) have been reported [31] and [32]. In this review, we provide the readers with an overview of current methods utilized in these crystallographic studies to express, purify and crystallize MAP kinases and the MAP kinase pathway module components in the presence and absence of binding partners such as docking peptide and inhibitors.

## 2. Crystallization of uncomplexed MAP kinases

### 2.1. Protein constructs and use of tags

MAP kinases are 42 kDa proteins with relatively short extensions from the protein kinase core at the N- and C-termini, and are known to fold into a compact

structure with the extensions packed on the kinase core. Thus, most crystallization experiments have involved the full-length MAP kinases (Table 1). In general, the full-length MAP kinases have been expressed with an N-terminal His$_6$-tag to facilitate protein purification. Unlike the vast majority of purification schemes used for MAP kinases, however, JNK3 was produced with a modified expression vector pET-15b in which the sequence encoding the N-terminal His$_6$-tag and thrombin cleavage site was eliminated by PCR [22]. Also, truncations in both N- and C-termini were necessary to crystallize JNK3 (JNK3/ΔN39/ΔC20). In the case of JNK1, only C-terminal truncation (Δ20) was made to crystallize the target protein [27]. The exceptional cases in terms of tag location include JNK1 (MAPK) [27] and MEK1/2 (MAP2K) [31]. Both JNK1 and MEK1/2 were expressed with a C-terminal His$_6$-tag. The His$_6$-tag is either directly fused to the coding sequence or with intervening linker sequence encoding a protease cleavage site. For example, the expression vector for ERK2 was NpT7-5 [23], [33] and [34] in which the His$_6$-tag is directly adjacent to the inserted gene. In contrast, p38α was expressed in the vector pET-14b [20], which encodes a His$_6$-tag and thrombin cleavage site, and several studies employed tags encoding a TEV protease cleavage site [28] and [29].

## 2.2. Expression and purification of low activity MAP kinases

### 2.2.1. Bacterial and Baculovirus/Insect cell expression systems

Bacterial expression is most convenient when adequate expression levels (5 mg/L or greater) are available. Most of low activity MAP kinases crystallized were expressed from *Escherichia coli* (Table 1). Bacterial expression has the added advantage of low levels of unwanted phosphorylation. However, some mammalian kinases do not express well in bacteria. The only other system that has been used in MAP kinase crystallization studies has been the Baculovirus/Insect cell system. Human p38 CSBP2 was expressed from Hi5 insect cell and crystallized in its apo, unphosphorylated form [21].

Most protocols for purifying MAP kinases for crystallization employ $Ni^{2+}$ NTA–agarose, making use of the incorporated $His_6$-tag. In each study, the protein was eluted with imidazole in the presence of NaCl, which reduces binding of untagged protein impurities. Both step and gradient elution have been employed [35] and [36]. No reducing agent is used in the $Ni^{2+}$ NTA elution to avoid reduction of the $Ni^{2+}$ cation on the column. It should be mentioned that, in the case of JNK3, $Ni^{2+}$ NTA–agarose was not employed, owing to the absence of $His_6$-tag in the modified expression vector, as discussed above [22]. Often, a second or third column is required to obtain sufficient purity for crystallization. The methods used have been hydrophobic interaction, size exclusion, cation exchange, or chromatofocussing (MonoP) chromatography. Low activity ERK2, however, does not require high purity for crystallization, for which purification with $Ni^{2+}$–NTA

followed by anion exchange is sufficient. Our experience is that reducing the overall timeframe for purification improves the likelihood of obtaining crystals as well as crystal quality.

## 2.3. Expression and purification of active, phosphorylated MAP kinases

Thus far, several structures of active, phosphorylated MAP kinases and MAP kinase pathway components have been reported, solved by X-ray crystallography. A variety of strategies has been utilized to make phosphorylated form of active kinases for crystallographic studies. The simplest is expression in the Baculovirus/Insect cell system, which can lead to constitutively phosphorylated kinases without other intervention. On the other hand, MAP kinases have been coexpressed with activator in bacteria [33] and [34] and insect cells. Also, some studies have utilized in vitro kinase activation.

### 2.3.1. Baculovirus/Insect cell system

The structure of the MAP3K Tao2 kinase domain (1–320) was solved in its phosphorylated active conformation by utilizing Baculovirus/Insect cell system [25]. For the crystallographic studies of this active from of protein kinase, the rat TAO2 kinase domain gene was cloned into pRSETB to incorporate N-terminal

His$_6$-tag and transferred into the baculovirus shuttle vector pVL1393. Selected recombinant viruses were replicated in Sf9 or Hi5 insect cells [37]. Also, the TAO2 kinase domain containing seleno-methionine was produced successfully in Hi5 insect cells by growing cells in standard medium, followed by l-methionine depletion, and replacement with selenomethione-supplemented medium as described [38].

### 2.3.2. Coexpression of target kinase with activating enzyme

The coexpression of target kinase with its upstream activating kinase is another way of producing active, phosphorylated kinases for crystallization. Coexpression of two genes in bacteria can be accomplished with either a one plasmid or two plasmid system. In the one plasmid method, two genes of interest can be inserted into dual expression vector such as pET Duet vectors from Novagen or NpT7-5 [33] and [34]. The crystal structure of doubly phosphorylated ERK2 was determined by utilizing the one-plasmid method [23]. In this study, the ERK2 was coexpressed with a constitutively active MEK1 (R4F) [39] and activated in vivo. For this one-plasmid coexpression system, both ERK2 and R4F were subcloned into the same NpT7-5 plasmid used to transform BL21(DE3) *E. coli* cells [33] and [34]. This active, doubly phosphorylated form of ERK2 was separated from the different phosphorylation states of ERK2 during anion-exchange step. The Baculovirus/Insect cell system is also available for the coexpression [40] and [41],

but the use of this method for making active MAP kinases or MAP kinase module components has not been reported.

### 2.3.3. In vitro kinase activation

The structure of doubly phosphorylated p38γ was determined by X-ray crystallography by using in vitro kinase activation [24]. The purified p38γ was incubated at room temperature with constitutively active MAP kinase kinase 6 (MKK6) in the presence of ATP and $MgCl_2$. The state of p38γ phosphorylation was confirmed by electrospray mass spectrometry. In a second study, non-phosphorylated Fus3 was incubated with pepSte5 as an activator in the presence of $MgCl_2$ overnight at room temperature to produce tyrosine phosphorylated Fus3 [29]. This reaction mixture was then separated on the gel filtration column to isolate the phosphorylated target kinase Fus3 from the activator pepSte5. In general, active, phosphorylated form of target kinase must be purified away from unphosphorylated and partially phosphorylated forms. This purification has been accomplished using anion exchange and/or hydrophobic interaction chromatography [23].

## 2.4. Crystallization

The original crystallization conditions for MAP kinases were found by methods standard for soluble proteins. The protein is concentrated until it starts to

precipitate, generally in neutrally buffered solution, at about 0.1 M NaCl, to determine the solubility. Then hanging drop vapor diffusion is set up below the solubility limit. Sparse matrix screens are commercially available from Hampton (Laguna Miguel, CA) and Emerald Biostructures (Bainbridge Island, WA). Crystallization conditions for uncomplexed MAP kinases are given in Table 1. There are many similarities among these conditions. In our work, we normally held the protein concentration near 8 mg/ml although the higher concentrations can be used. The pH was nearly always maintained in the neutral range. However, in the crystallization of ERK2, the precipitant solutions were slightly acidic. Note that the buffer concentration is often greater in the precipitant solution, thus introducing a change in pH in the droplet.

### 2.4.1. Truncations

As in crystallizations of many proteins, a variety of strategies has been employed to facilitate MAP kinase crystal growth, one of which is construct modification [42].

The crystallization of human JNK3α1 was improved by making both N-terminal and C-terminal truncations (JNK3/ΔN39/ΔC20) [22]. Structure-based sequence alignment of human JNK3α1 with p38 and ERK2 reveals that it is longer than the others at both the N- and C-termini. Deletion of both extensions was necessary to

obtain crystals. MAP3K TAO2 is 1241 residues in length, but the full length protein does not express well. On the other hand, the kinase domain alone (TAO2/1−320) does express in insect cell and crystallizes [25].

## 2.4.2. Surface engineering by mutagenesis

Mutation of surface residues has been used several times to improve protein crystals. The approaches are to eliminate hydrophobic patches, or to introduce charge changes [42] and [43]. A recent study revealed that crystals of p38α could be improved by eliminating a cysteine residue [35]. The authors observed that p38α oligomerized in non-reducing solutions, and found that C162 was oxidized. A mutant p38α/C162S displayed the enhanced solubility and homogeneity. Crystals were easier to obtain, revealed diffraction to 2.0 Å, and had lower mosaicity [35]. The mutant p38α/C162S showed only slightly reduced enzymatic activity. This newly identified mutant, p38α/C162S, could become a standard reagent for the studies inhibitors of p38α inhibitors.

## 2.4.3. Improved expression systems

A common bottleneck of protein crystallization generally is poor expression of the protein of interest in bacteria. There are probably undiscovered reasons for this, but part of the problem is known to be differences in codon usage between bacteria and the introduced cDNA. The modified *E. coli* strain Rosetta (DE3) is

optimized for expression of eukaryotic proteins translated by the codons rarely found in *E. coli*. Bukhtiyarova et al. [44] used Rosetta (DE3) to make p38α, and were able to produce over 70 mg/L of bacterial culture. This group also reported other improvements to the purification and crystallization of p38α, including a shortened purification protocol that could be completed within 36 h, reducing oxidation and degradation. Another approach to the problem of codon usage is to make synthetic genes, utilizing bacterial codons [45], although this technique has not been reported for MAP kinases or MAP kinase pathway components so far.

## 3. Crystallization of MAP kinases with docking peptide

Through studies in many laboratories, it has been demonstrated that MAP kinases bind substrates, scaffolds, activating MAP2Ks and phosphatases through docking interactions [4], [5] and [7]. Crystallographic analyses of MAPKs complexed with peptides derived from these interacting partners are elucidating the structural basis of MAP kinase docking interactions and docking-interaction-induced conformational changes [26], [27], [28], [29] and [30] (Table 2). Docking motifs in substrates and other interacting partners are known to be short linear sequences spanning between 11 and 18 residues in length. Thus it is relatively straightforward to synthesize peptides corresponding to the docking motif and carry out co-crystallization trials. The reported crystallizations have many similarities. Prior to its crystallization, the protein:peptide complex was formed by

mixing protein kinase with chemically synthesized peptide in 1:2–5 molar ratio. All of MAP kinases crystallized in complex with docking peptides utilized PEG as main precipitant. Further, in most cases, crystals were obtained by the vapor diffusion hanging drop method. On the other hand, several unique protocols have been reported to improve crystallization of MAP kinases bound with docking peptides.

## 3.1. Various lengths of docking peptide

The length of docking peptide employed was varied dramatically in the study of Fus3 bound by docking motifs derived from Ste5, up to 30 residue [29], and quite long peptides were also employed in a Fus3/Msg5-peptide complex [28]. The authors noted that the length of docking peptide sequence affected the crystallization. Our experience was that the exact identity of the peptide had a great effect on both the diffraction limit and the electron density for the bound peptide, and that good results could only be obtained by screening numerous peptides in co-crystallization experiments.

## 3.2. Use of tethering

The recent crystal structure of unphosphorylated ERK2 complexed with 16-mer (V31C) D-motif pepHePTP shows an example in which artificial generation of tethering between docking peptide and docking groove through a disulfide bond

can be employed as a useful tool for improving crystallization of MAP kinase:peptide complex [69] (Table 2). In this study, the initial crystal of inactive ERK2 in complex with HePTP-derived peptide containing the sequence $^{16}$RLQERRGSNVALMLDV$^{31}$ diffracted to 2.9 Å. The mutant ERK2/T116C and a peptide with the sequence $^{16}$RLQERRGSNVALMLDC$^{31}$ were generated to induce the formation of disulfide bond as observed in p38α-pepMKK3b complex. The complex of ERK2/T116C and pepHePTP (V31C) was crystallized under the same condition but diffracted to 1.9 Å.

## 3.3. Use of phosphorylation site mutants and inhibitors

In the crystallization of low activity, unphosphorylated kinase, introducing mutations on the phosphorylation sites of the activation loop has been employed as a useful way of producing highly homogeneous target protein. This approach was also used to crystallize yeast homolog MAP kinase Fus3 complexed with docking peptides [28] and [29]. In this study, the double phosphorylation sites on the activation loop were mutated into valine and phenylalanine to make this kinase non-phosphorylatable for its enhanced crystallization. The structure of human JNK1 complexed with a scaffold-derived peptide (JIP1) was crystallized with an inhibitor bound at the active site [27] (Table 2).

## 4. Crystallization of MAP kinases with drug-like inhibitors

Crystallography has featured prominently in development of inhibitors for MAP kinase module components. A substantial body of work, primarily from the drug industry, has been reported in over 20 papers, with about 50 crystal structures on deposit at the Protein Data Bank (Table 3). The crystal structures of MAP kinases complexed with inhibitors have helped to elucidate the structural basis of inhibitor specificity and potency [46], [47] and [48] and is used in lead optimization [49], [50] and [51]. Although many inhibitors are ATP competitive, structural studies have revealed different binding sites and different mechanisms of kinase inhibition in a few cases [31] and [52]. Recently, X-ray crystallography is being used in lead generation using low-molecular weight "fragments" [49], [50] and [51] introduced by soaking. Here, we present the current methods to facilitate the MAP kinase crystallization complexed with its inhibitors.

## 4.1. Co-crystallization versus soaking

There are two ways of forming crystals of protein–inhibitor complexes, one is to soak preexisting crystals in the compound of interest, and the other is to co-crystallize, setting up the crystallization trial in the presence of inhibitor. About half of reported structures of MAP kinase–inhibitor complexes are from crystals formed by co-crystallization (Table 3). The protein sample is preincubated with a

3- to 10-fold molar excess of inhibitor before crystallization. In the soaking method, the crystal is formed first and then the compound is added, often dissolved in DMSO, to crystallization droplet such that the final concentration is in 10-fold excess over the protein present, or as much as 0.1 mM.

## 4.2. Crystallization conditions

Conditions used to obtain complexes of MAP kinases with inhibitors tend to be surprisingly similar. Nearly all of the reported crystallization conditions involve the use of polyethylene glycols (PEGs), with medium molecular weight PEGs predominating. It is possible that PEGs do not interfere or compete strongly with the inhibitors as might be the case with either more hydrophobic compounds or with salts. Low amounts (0.2 M) of either $Mg(OAc)_2$ or $(NH_4)_2SO_4$ are also prevalent in the crystallization conditions, a common feature of conditions available in commercial kits [53].

## 4.3. Truncations

Most often, full-length MAP kinases have been used for co-crystallization with inhibitors. However, there are examples in which the truncation of target protein was necessary for crystallization. Truncations were used in JNK as noted above [27] and [48]. MEKs 1 and 2 studied in complex with inhibitors were particularly difficult to crystallize. Ultimately, the unphosphorylated human MEK1 and

MEK2 proteins were truncated at the N-terminus to facilitate crystallization by 61 and 54 residues, respectively [31], after trying several N-terminal truncations. Also, truncated B-RAF (MAP3K) kinase domain was crystallized as described [32]. A His$_6$-tag was used in MEKs 1 and 2, but the tag was fused to C-terminus, unlike the most cases of MAP kinases crystallized.

## 5. Concluding remarks

The recent years have witnessed noticeable growth in information relating to MAP kinase pathway regulatory and specificity-determining mechanisms. Crystallographic studies have played an important role in defining these mechanisms, owing to a truly remarkable success in crystallizing MAP kinases. This success may be due to their small size, and the ability to express them in *E. coli*. MAP2Ks and MAP3Ks have not proved to crystallize easily so far, and have received less attention. However, with advances in our understanding of what constitutes well-folded domains of MAP2Ks and MAP3Ks, how they interact with each other and other interacting partners, the ability to make numerous constructs, and the use of synthetic genes, should improve our progress in structural studies of these proteins as well in the future.

Table 1
Crystallization of uncomplexed MAP kinases

| MAP kinases | | Expression system | Crystallization condition | Space group/unit cell dimensions | Resolution (Å) | Ref. |
|---|---|---|---|---|---|---|
| MAPK (inactive) | Rat ERK2 | E. coli | 18% (w/v) PEG 8000, 0.2 M $(NH_4)_2SO_4$, 5 mM DTT, 0.1 mM $NaN_3$, 50 mM MES, pH 5.9 | $P2_1/a = 49.32$ Å, $b = 71.42$ Å, $c = 61.25$ Å, $\beta = 109.75°$ | 2.3 | [19] |
| | Rat ERK2 (K52R/MgATP) | E. coli | 18% (w/v) PEG 8000, 0.2 M $(NH_4)_2SO_4$, 5 mM DTT, 0.1 mM $NaN_3$, 50 mM MES, pH 5.9 | $P2_1/a = 49.32$ Å, $b = 71.42$ Å, $c = 61.25$ Å, $\beta = 109.75°$ | 2 | [36] |
| | Human p38 CSBP2 | Insect cell | 0.56 M sodium citrate 0.13 M $(NH_4)_2SO_4$, 40 mM Na–Hepes pH 7.5 | $P2_12_12_1/a = 45.2$ Å, $b = 86.21$ Å, $c = 123.9$ Å | 2.3 | [21] |
| | Mouse p38α | E. coli | 18% (w/v) PEG 8000, 0.2 M $(NH_4)_2SO_4$, 0.1 M Hepes, pH 7.5 | $P2_12_12_1/a = 45.76$ Å, $b = 84.93$ Å, $c = 123.91$ Å | 2.1 | [20] |
| | Human JNK3α1 (ΔN39, ΔC20/MgAMP-PNP) | E. coli | 18–20% (v/v) PEG 550, 10% (v/v) ethylene glycol, 20 mM β-mercaptoethanol, 100 mM Hepes, pH 7.5 | $P2_12_12_1/a = 51.50$ Å, $b = 71.24$ Å, $c = 107.60$ Å | 2.3 | [22] |
| | Yeast Fus3 (w ADP) | E. coli | 25–28% PEG 1000, 5–10% MPD, 0.1 M MES, pH 6.1 | $P2_12_12_1/a = 56.7$ Å, $b = 62.5$ Å, $c = 87.17$ Å | 1.8 | [28] |
| | Yeast Fus3 (VF)[a] (MgATP-γS) | E. coli | 18% PEG 3350, 10% MPD, 0.2 M KSCN, 0.1 M MES, pH 6.1 | $P2_12_12_1/a = 79.15$ Å, $b = 94.47$ Å, $c = 113.2$ Å | 2.85 | [29] |
| MAPK (active)[b]  T*EY* | Rat ERK2 | E. coli | 20% (w/v) PEG 4000, 0.2 M $(NH_4)_2SO_4$, 0.1 M MES, pH 6.5 | $P4_12_12/a = b = 92.5$ Å, $c = 103.0$ Å | 2.4 | [23] |
| T*GY* | Human p38γ (MgAMP-PNP) | E. coli | 27% PEG 4000, 0.1 M NaOAc 10 mM $MgCl_2$, 0.1 M Tris, pH 8.5, 5 mM DTT | $P2_12_12/a = 63.50$ Å, $b = 66.82$ Å, $c = 206.02$ Å | 2.4 | [24] |
| T*GY* | Human p38α (carboxy-methylated) | E. coli | 0.62–0.69 M sodium citrate 0.13 M $(NH_4)_2SO_4$, 40 mM Hepes, pH 7.5 | $P2_12_12/a = 45.2$ Å, $b = 84.19$ Å, $c = 120.04$ Å | 2.3 | [35] |
| T*GY* | Human p38α (C162S) | E. coli | 0.62–0.69 M sodium citrate, 0.13 M $(NH_4)_2SO_4$, 40 mM Hepes, pH 7.5 | $P2_12_12/a = 45.2$ Å, $b = 84.80$ Å, $c = 124.91$ Å | 2.0 | [35] |
| TEY* | Yeast Fus3 (MgATP-γS) | E. coli | 25–28% PEG 1000, 5–10% MPD, 0.1 M MES, pH 7.5 | $P2_12_12/a = 56.8$ Å, $b = 62.5$ Å, $c = 86.0$ Å | 2.1 | [29] |
| MAP3K (active)  ANS*FVG | Rat TAO2 (phosphorylated/Apo form) | Insect cell | 18–20% PEG 1000, 0.2 M $Ca(OAc)_2$, 0.1 M imidazole, pH 7.8–8.2 | $P6_222/a = b = 186.22$ Å, $c = 94.51$ Å | 2.1 | [25] |
| ANS*FVG | Rat TAO2 (phosphorylated/MgATP) | Insect cell | 18–20% PEG 1000, 0.2 M $Ca(OAc)_2$, 0.1 M imidazole, pH 7.8–8.2 | $P6_222/a = b = 186.47$ Å, $c = 91.10$ Å | 2.1 | [25] |

[a] Thr180 and Tyr182 were mutated to Val and Phe, respectively.

[b] Asterisk (*) indicates the phosphorylation of the noted residue.

Table 2
Crystallization of MAP kinases with docking peptide

| MAP kinases | | Docking peptide (sequence) | Crystallization condition | Space group/unit cell dimensions | Resolution (Å) | Ref. |
|---|---|---|---|---|---|---|
| MAPK | Mouse p38α | pepMEF2A[a] (RKPDLRVVIPPS) | 20% (w/v) PEG 8000, 0.1 M Hepes, pH 7.5 | $P3_121/a = 81.9$ Å, $c = 123.0$ Å | 2.3 | [26] |
| | Mouse p38α | pepMKK3b[b] (SKGKSKRKKDLRISCNSK) | 20% (w/v) PEG 8000, 0.2 M calcium acetate, 0.1 M sodium cacodylate, pH 7.0 | $P3_121/a = 82.3$ Å, $c = 123.3$ Å | 2.3 | [26] |
| | Human JNK1α1 (ΔC20) | pepJIP1[c] (RPKRPTTLNLF) | 18–22% (v/v) PEG 3000, 0.1 M tri-sodium citrate pH 5.5 | $P2_12_12_1/$ $a = 62.40$ Å, $b = 80.84$ Å, $c = 83.74$Å | 2.35 | [27] |
| | Human JNK1α1 (ΔC20) | pepJIP1/SP600125[d] (RPKRPTTLNLF) | 18–22% (v/v) PEG 3000, 0.1 M tri-sodium citrate pH 5.5 | $P2_12_12_1/$ $a = 61.69$ Å, $b = 79.46$ Å, $c = 82.86$ Å | 2.7 | [27] |
| | Yeast Fus3 (VF)[e] (MgATP-γS) | pepSte7[f] (RRNLKGLNLNLHPD) | 15–20% PEG 8000, 0.1 M MES, pH 6.1 | $P2_12_12_1/$ $a = 58.3$ Å, $b = 63.8$ Å, $c = 99.6$ Å | 1.55 | [28] |
| | Yeast Fus3 (VF)[e] (MgATP-γS) | pepMsg5[g] (PRSLQNRNTKNL SLDIAALHP) | 15–20% PEG 8000, 0.1 M MES, pH 6.1 | $P2_12_12_1/$ $a = 58.1$ Å, $b = 63.62$ Å, $c = 99.95$ Å | 2.5 | [28] |
| | Yeast Fus3 (VF)[e] (MgATP-γS) | pepFar1[h] (SKRGNIPKPLNLS) | 15–20% PEG 8000, 0.1 M MES, pH 6.1 | $P2_12_12_1/$ $a = 58.3$ Å, $b = 63.24$ Å, $c = 100.85$ Å | 2.3 | [28] |
| | Yeast Fus3 (VF)[e] (MgATP-γS) | pepSte5[i] (TPVERQTIYSQA PSLNPN LILAAPPKERNQ) | 18% PEG 3350, 10% MPD, 0.2 M KSCN, 0.1 M MES, pH 6.1 | $P2_12_12_1/$ $a = 94.62$ Å, $b = 95.21$ Å, $c = 101.66$ Å | 1.9 | [29] |
| | Rat ERK2 | pepHePTP[j] ([16]RLQERRGSNVALMLDV[31]) | 20% PEG 4000, 10% isopropanol, 0.1 M Na–Hepes, pH 7.5 | $P2_12_12_1/$ $a = 45.9$ Å, $b = 66.0$ Å, $c = 116.4$ Å | 2.9 | [69] |
| | Rat ERK2 (T116C) | pepHePTPm(V31C) ([16]RLQERRGSNV ALMLDC[31]) | 20% PEG 4000, 10% isopropanol, 0.1 M Na–Hepes pH 7.5 | $P2_12_12_1/$ $a = 46.0$ Å, $b = 66.3$ Å, $c = 116.1$ Å | 1.9 | [69] |
| | Rat ERK2 | pepMKP3[k] ([60]GIMLRRLQKG NLPVRAL[76]) | 20% PEG 3350, 0.1–0.2 M (NH₄)Cl, 0.1 M 2-morpholino-ethanesulfonic acid, pH 6.5 | $P2_1/a = 57.4$ Å, $b = 67.5$ Å, $c = 86.6$ Å, $β = 99.6°$ | 2.5 | [30] |

[a] pepMEF2A is derived from p38 transcription factor substrate MEF2A.
[b] pepMKK3b is derived from p38 activating enzyme MKK3b.
[c] pepJIP1 is 11mer peptide derived from JNK-interacting protein-1, which is a scaffolding protein assembling the components of the JNK cascade.
[d] SP600125 is a reversible ATP-competitive inhibitor of JNK.
[e] Thr180 and Tyr182 in the activation segment were mutated to Val and Phe, respectively. Thus, Fus3 (VF) is non-phosphorylatable.
[f] pepSte7 is derived from Fus3-activating enzyme Ste7 (MAPKK).
[g] pepMsg5 is derived from Fus3-inactivating phosphatase Msg5.
[h] pepFar1 is derived from Fus3 substrate Far1.
[i] pepSte5 is derived from the scaffold Ste5.
[j] pepHePTP is derived from hematopoietic protein tyrosine phosphatase.
[k] pepMKP3 is derived from MAP kinase phosphatase.

Table 3
Crystallization of MAP kinases with drug-like inhibitor

| MAP kinases | Inhibitor | Crystallization condition | Space group/unit cell dimensions | Resolution (Å) | Ref. |
|---|---|---|---|---|---|
| MAPK | | | | | |
| Human p38 (CSBP2) | 3'-Iodo pyridinyl-imidazole | 28-32% (w/v) PEG 1500, 5 mM DTT (co-crystallization) | $P2_12_12_1/a = 65.1$ Å, $b = 74.3$ Å, $c = 77.9$ Å | 2.0 | [54,55] |
| Mouse p38 | SB 203580 | 18% (w/v) PEG 8000, 0.2 M Mg(OAc)$_2$, 0.1 M Hepes, pH 7.5 (soaking crystal) | $P2_12_12_1/a = 45.48$ Å, $b = 85.03$ Å, $c = 123.13$ Å | 2.4 | [46] |
| Mouse p38 | SB 216995 | 18% (w/v) PEG 8000, 0.2 M Mg(OAc)$_2$, 0.1 M Hepes, pH 7.5 (soaking crystal) | $P2_12_12_1/a = 45.47$ Å, $b = 84.71$ Å, $c = 123.54$ Å | 2.5 | [46] |
| Mouse p38 | SB 218655 | 18% (w/v) PEG 8000, 0.2 M Mg(OAc)$_2$, 0.1 M Hepes, pH 7.5 (soaking crystal) | $P2_12_12_1/a = 45.58$ Å, $b = 85.22$ Å, $c = 123.50$ Å | 2.4 | [46] |
| Mouse p38 | SB 220025 | 18% (w/v) PEG 8000, 0.2 M Mg(OAc)$_2$, 0.1 M Hepes, pH 7.5 (soaking crystal) | $P2_12_12_1/a = 45.50$ Å, $b = 84.83$ Å, $c = 123.67$ Å | 2.5 | [46] |
| Rat ERK2 | SB 220025 | 18% (w/v) PEG 8000, 0.2 M (NH$_4$)$_2$SO$_4$, 5 mM DTT, 0.1 mM NaN$_3$, 50 mM MES, pH 5.9 (soaking crystal) | $P2_1/a = 48.85$ Å, $b = 70.0$ Å, $c = 60.70$ Å, $\beta = 109.20°$ | 2.1 | [46] |
| Rat ERK2 | Olomoucine | 18% (w/v) PEG 8000, 0.2 M (NH$_4$)$_2$SO$_4$, 5 mM DTT, 0.1 mM NaN$_3$, 50 mM MES, pH 5.9 (soaking crystal) | $P2_1/a = 48.87$ Å, $b = 70.26$ Å, $c = 59.97$ Å $\beta = 109.05°$ | 2.1 | [46] |
| Human ERK2 (I103L, Q105T, D106H, E109G, T110A) | SB 203580 | 28-30% (v/v) PEG 2000, 0.2 M (NH$_4$)$_2$SO$_4$, 20 mM β-mercaptoethanol, 0.1 M Hepes, pH 7.2 (soaking crystal) | $P2_1/a = 48.64$ Å, $b = 69.68$ Å, $c = 60.28$ Å $\beta = 109.25°$ | 2.0 | [56] |
| Human p38 | Compound 3 (quinazoline) | 30% (v/v) PEG 600, 0.25 M (NH$_4$)$_2$SO$_4$, 0.1 M Hepes, pH 7.5 (co-crystallization) | $P2_12_12_1/a = 45.18$ Å, $b = 84.86$ Å, $c = 127.01$ Å | 2.6 | [57] |
| Human p38 | Compound 1 (diaryl urea) BIRB 796 | 28-32% (w/v) PEG 1500, 2-5 mM TFA 5 mM DTT (co-crystallization) | $P2_12_12_1/a = 64.00$ Å, $b = 74.00$ Å, $c = 76.80$ Å | 2.8 | [52] |
| Mouse p38α | Compound 1 (pyrimidine imidazole) | 0.56 M sodium citrate, 0.13 M (NH$_4$)$_2$SO$_4$, 40 mM Na-Hepes, pH 7.5 (soaking crystal) | $P2_12_12_1/a = 45.70$ Å, $b = 88.04$ Å, $c = 126.05$ Å | 2.5 | [47] |
| Mouse p38α | Compound 3 (quinazolinone) | 0.56 M sodium citrate, 0.13 M (NH$_4$)$_2$SO$_4$, 40 mM Na-Hepes, pH 7.5 (soaking crystal) | $P2_12_12_1/a = 45.01$ Å, $b = 87.10$ Å, $c = 122.61$ Å | 2.1 | [47] |
| Mouse p38α | Compound 4 (pyridol-pyrimidine) | 0.56 M sodium citrate, 0.13 M (NH$_4$)$_2$SO$_4$, 40 mM Na-HEPES, pH 7.5 (soaking crystal) | $P2_12_12_1/a = 45.63$ Å, $b = 86.76$ Å, $c = 126.40$ Å | 2.5 | [47] |
| Human p38 | 4-Azaindole inhibitor | 17% PEG 3350, 50 mM CaCl$_2$, 50 mM Hepes, pH 7.6 (soaking crystal) | $P2_12_12_1/a = 45.40$ Å, $b = 85.45$ Å, $c = 125.84$ Å | 2.1 | [58] |
| Human p38 | Compound 14e (dihydro-quinazolinone) | Sodium citrate, (NH$_4$)$_2$SO$_4$, Hepes pH 7.0 (soaking crystal) | $P2_12_12_1/a = 45.52$ Å, $b = 87.16$ Å, $c = 124.94$ Å | 2.4 | [59] |
| Human ERK2 | Compound 3 (pyridazine derivative) | 35% (v/v) PEG 5000, 0.2 M (NH$_4$)$_2$SO$_4$, 0.1 M Bis-Tris, pH 6.5 (co-crystallization) | $P2_1/a = 48.60$ Å, $b = 68.54$ Å, $c = 59.83$ Å $\beta = 110.3°$ | 2.5 | [60] |
| Human p38 | Compound 39 | Sodium citrate, (NH$_4$)$_2$SO$_4$, Hepes pH 7.5 (no detail) | $P2_12_12_1/a = 45.57$ Å, $b = 87.47$ Å, $c = 122.64$ Å | 2.0 | [61] |
| Human JNK3α1 (ΔN39, ΔC20/MgAMP-PCP) | Compound 1 (imidazole-pyrimidine) | 20% PEG MME 550, 10% ethylene glycol, 0.1 M Hepes pH 7.3 (soaking crystal) | $P2_12_12_1/a = 53.07$ Å, $b = 71.34$ Å, $c = 107.17$ Å | 2.2 | [48] |
| Human JNK3α1 (ΔN39, ΔC20/MgAMP-PCP) | Compound 2 (imidazole-pyrimidine) | 20% PEG MME 550 10% ethylene glycol 0.1M HEPES pH 7.3 (soaking crystal) | $P2_12_12_1/a = 50.95$ Å, $b = 71.53$ Å, $c = 106.00$ Å | 2.2 | [48] |
| Human JNK3α1 (ΔN39, ΔC20/MgAMP-PCP) | Compound 3 (phenanthroline) | 20% PEG MME 550, 10% ethylene glycol, 0.1 M Hepes, pH 7.3 (co-crystallization) | $P2_12_12_1/a = 62.40$ Å, $b = 62.38$ Å, $c = 98.12$ Å | 2.7 | [48] |

Table 3 (continued)

| MAP kinases | Inhibitor | Crystallization condition | Space group/unit cell dimensions | Resolution (Å) | Ref. |
|---|---|---|---|---|---|
| Human JNK3γ1 (ΔN39, ΔC20/MgAMP-PCP) | Compound 4 (dihydro-anthrapyrazole) | 20% PEG MME 550, 10% ethylene glycol, 0.1 M Hepes pH 7.3 (soaking crystal) | $P2_12_12_1/a = 48.21$ Å, $b = 73.09$ Å, $c = 106.04$ Å | 2.5 | [48] |
| Human ERK2 | FR180204 | 30–35% PEG 5000, 5 mM DTT, 0.15 M (NH₄)₂SO₄, 0.1 M Bis-Tris-HCl, pH 6.5 (co-crystallization) | $P2_1/a = 48.86$ Å, $b = 69.99$ Å, $c = 63.30$ Å, $\beta = 116.5°$ | 2.5 | [62] |
| Mouse p38α | Compound 1 | 18% (w/v) PEG 8000, 0.2 M MgOAc)₂, 0.1 M Hepes, pH 7.5 (soaking crystal) | $P2_12_12_1/a = 45.77$ Å, $b = 87.10$ Å, $c = 126.34$ Å | 2.21 | [51] |
| Mouse p38α | Compound 2 | 18% (w/v) PEG 8000, 0.2 M MgOAc)₂, 0.1 M Hepes, pH 7.5 (soaking crystal) | $P2_12_12_1/a = 45.75$ Å, $b = 84.72$ Å, $c = 126.76$ Å | 2.16 | [51] |
| Mouse p38 (S180A, Y182F) | Compound 3 (isoxazolone) | PEG 1500, pH 7.0 (co-crystallization) | $P2_12_12_1/a = 65.04$ Å, $b = 76.95$ Å, $c = 73.64$ Å | 2.01 | [63] |
| Mouse p38α (S180A, Y182F) | Compound 15d (monocyclic pyrazolone) | 12–17% PEG 4000, 0.1 M MES, pH 6.5 (co-crystallization) | $P2_12_12_1/a = 65.74$ Å, $b = 74.57$ Å, $c = 77.38$ Å | 1.95 | [64] |
| Human p38α | Inhibitor 1a | 10–20% PEG 4000, 50 mM $n$-octyl-β-D-glucoside, 0.1 M cacodylic acid, pH 6.0 (co-crystallization) | $P2_12_12_1/a = 67.76$ Å, $b = 70.92$ Å, $c = 75.80$ Å | 2.0 | [65] |
| Human p38α | Inhibitor 2a | 10–20% PEG 4000, 50 mM $n$-octyl-β-D-glucoside, 0.1 M cacodylic acid, pH 6.0 (co-crystallization) | $P2_12_12_1/a = 64.29$ Å, $b = 74.23$ Å, $c = 74.42$ Å | 2.0 | [65] |
| Human p38α | Compound 25 CP-808844 | Na–K tartrate, trimethylamine, Tris, pH 8.0 (no detail) | $P2_12_12_1/a = 45.46$ Å, $b = 86.63$ Å, $c = 122.86$ Å | 2.0 | [66] |
| Human p38α | Ro3201195 (triazolopyridine) | 6–10% (w/v) PEG 5000, cacodylic acid, pH 6.5 (co-crystallization) | $P2_12_12_1/a = 64.8$ Å, $b = 74.7$ Å, $c = 76.9$ Å | 2.8 | [67] |
| Human p38α | Pyrazoloamine | 6–10% (w/v) PEG 5000, cacodylic acid, pH 6.5 (co-crystallization) | $P2_12_12_1/a = 65.2$ Å, $b = 75.2$ Å, $c = 77.8$ Å | 2.1 | [67] |
| Human p38α | MPAQ | 6–10% (w/v) PEG 5000, cacodylic acid, pH 6.5 (co-crystallization) | $P2_12_12_1/a = 66.3$ Å, $b = 75.8$ Å, $c = 78.1$ Å | 2.2 | [67] |
| Human p38α | Pyrazolourea | 6–10% (w/v) PEG 5000, cacodylic acid, pH 6.5 (co-crystallization) | $P2_12_12_1/a = 59.0$ Å, $b = 67.1$ Å, $c = 87.7$ Å | 2.25 | [67] |
| Human p38α | Compound 1 (pyridine) | 17% (w/v) PEG 5000, 0.2 M (NH₄)₂SO₄, 0.1 M Hepes, pH 7.0 (soaking crystal) | $P2_12_12_1/a = 45.77$ Å, $b = 87.10$ Å, $c = 126.34$ Å | 2.2 | [49] |
| Human p38α | Compound 2 (indole) | 17% (w/v) PEG 5000, 0.2 M (NH₄)₂SO₄, 0.1 M Hepes, pH 7.0 (soaking crystal) | $P2_12_12_1/a = 45.64$ Å, $b = 86.60$ Å, $c = 124.92$ Å | 2.2 | [49] |
| Human p38α | Compound 5 | 17% (w/v) PEG 5000, 0.2 M (NH₄)₂SO₄, 0.1 M Hepes, pH 7.0 (soaking crystal) | $P2_12_12_1/a = 45.3$ Å, $b = 85.91$ Å, $c = 125.71$ Å | 2.4 | [49] |
| Human p38α | Compound 9 | 17% (w/v) PEG 5000, 0.2 M (NH₄)₂SO₄, 0.1 M Hepes, pH 7.0 (soaking crystal) | $P2_12_12_1/a = 45.68$ Å, $b = 86.10$ Å, $c = 126.07$ Å | 2.5 | [49] |
| Human p38α | Compound 10 | 17% (w/v) PEG 5000, 0.2 M (NH₄)₂SO₄, 0.1 M Hepes, pH 7.0 (soaking crystal) | $P2_12_12_1/a = 46.14$ Å, $b = 86.00$ Å, $c = 125.73$ Å | 2.4 | [49] |
| Human p38α | Compound 16 | 17% (w/v) PEG 5000, 0.2 M (NH₄)₂SO₄, 0.1 M Hepes, pH 7.0 (soaking crystal) | $P2_12_12_1/a = 45.68$ Å, $b = 85.37$ Å, $c = 126.00$ Å | 2.2 | [49] |
| Human p38α | Compound 22 | 17% (w/v) PEG 5000, 0.2 M (NH₄)₂SO₄, 0.1 M Hepes, pH 7.0 (soaking crystal) | $P2_12_12_1/a = 45.5$ Å, $b = 86.86$ Å, $c = 125.20$ Å | 2.05 | [49] |
| Human p38α | Compound 23 | 17% (w/v) PEG 5000, 0.2 M (NH₄)₂SO₄, 0.1 M Hepes, pH 7.0 (soaking crystal) | $P2_12_12_1/a = 45.77$ Å, $b = 85.3$ Å, $c = 126.64$ Å | 1.97 | [49] |
| Human p38α | Compound 24 | 17% (w/v) PEG 5000, 0.2 M (NH₄)₂SO₄, 0.1 M Hepes, pH 7.0 (soaking crystal) | $P2_12_12_1/a = 45.35$ Å, $b = 85.46$ Å, $c = 125.60$ Å | 2.0 | [49] |

# References

[1] J.M. Kyriakis and J. Avruch, *Physiol. Rev.* **81** (2001), pp. 807–869.

[2] Z. Chen, T.B. Gibson, F. Robinson, L. Silvestro, G. Pearson, B. Xu, A. Wright, C. Vanderbilt and M.H. Cobb, *Chem. Rev.* **101** (2001), pp. 2449–2476.

[3] G. Pearson, F. Robinson, T. Beers Gibson, B.-e. Xu, M. Karandikar, K. Berman and M.H. Cobb, *Endocr. Rev.* **22** (2001), pp. 153–183.

[4] A.D. Sharrocks, S.-H. Yang and A. Galanis, *Trends Biochem. Sci.* **25** (2000), pp. 448–453.

[5] T. Tanoue and E. Nishida, *Cell. Signal.* **15** (2003), pp. 455–462.

[6] T. Tanoue, M. Adachi, T. Moriguchi and E. Nishida, *Nat. Cell Biol.* **2** (2000), pp. 110–116. View Record in Scopus | Cited By in Scopus (292)

[7] R.M. Biondi and A.R. Nebreda, *Biochem. J.* **372** (2003), pp. 1–13.

[8] J.M. English and M.H. Cobb, *Trends Pharmacol. Sci.* **23** (2002), pp. 40–45.

[9] J.S. Sebolt-Leopold, *Oncogene* **19** (2000), pp. 6594–6599.

[10] J.S. Sebolt-Leopold and R. Herrera, *Nat. Rev. Cancer* **4** (2004), pp. 937–947.

[11] J.C. Lee, J.T. Laydon, P.C. McDonnell, T.F. Gallagher, S. Kumar, D. Green, D. McNulty, M.J. Blumenthal, J.R. Keys, S.W. Land vatter, J.E. Strickler, M.M. McLaughlin, I.R. Siemens, S.M. Fisher, G.P. Livi, J.R. White, J.L. Adams and P.R. Young, *Nature* **372** (1994), pp. 739–746.

[12] K.M. Walton, R. DiRocco, B.A. Bartlett, E. Koury, V.R. Marcy, B. Jarvis, E.M. Schaefer and R.V. Bhat, *J. Neurochem.* **70** (1998), pp. 1764–1767.

[13] H. Schramek, *News Physiol. Sci.* **17** (2002), pp. 62–67.

[14] D. Bozyczko-Coyne, M.S. Saporito and R.L. Hudkins, *Curr. Drug Targets CNS Neurol. Disord.* **1** (2002), pp. 31–49.

[15] H. Davies, G.R. Bignell, C. Cox, P. Stephens, S. Edkins, S. Clegg, J. Teague, H. Woffendin, M.J. Garnett, W. Bottomley, N. Davis, E. Dicks, R. Ewing, Y. Floyd, K. Gray, S. Hall, R. Hawes, J. Hughes, V. Kosmidou, A. Menzies, C. Mould, A. Parker, C. Stevens, S. Watt, S. Hooper, R. Wilson, H. Jayatilake, B.A. Gusterson, C. Cooper, J. Shipley, D. Hargrave, K. Pritchard-Jones, N. Maitland, G. Chenevix-Trench, G.J. Riggins, D.D. Bigner, G. Palmieri, A. Cossu, A. Flanagan, A. Nicholson, J.W.C. Ho, S.Y. Leung, S.T. Yuen, B.L. Weber, H.F. Seigler, T.L. Darrow, H. Paterson, R. Marais, C.J. Marshall, R. Wooster, M.R. Stratton and P.A. Futreal, *Nature* **417** (2002), pp. 949–954.

[16] H. Rajagopalan, A. Bardelli, C. Lengauer, K.W. Kinzler, B. Vogelstein and V.E. Velculescu, Nature 418 (2002) 934–934.

[17] G. Singer, R. Oldt III, Y. Cohen, B.G. Wang, D. Sidransky, R.J. Kurman and I.-M. Shih, *J. Natl. Cancer Inst.* **95** (2003), pp. 484–486.

[18] Y. Cohen, M. Xing, E. Mambo, Z. Guo, G. Wu, B. Trink, U. Beller, W.H. Westra, P.W. Ladenson and D. Sidransky, *J. Natl. Cancer Inst.* **95** (2003), pp. 625–627.

[19] F. Zhang, A. Strand, D. Robbins, M.H. Cobb and E.J. Goldsmith, *Nature* **367** (1994), pp. 704–711.

[20] Z. Wang, P.C. Harkins, R.J. Ulevitch, J. Han, M.H. Cobb and E.J. Goldsmith, *Proc. Natl. Acad. Sci. USA* **94** (1997), pp. 2327–2332.

[21] K.P. Wilson, M.J. Fitzgibbon, P.R. Caron, J.P. Griffith, W. Chen, P.G. McCaffrey, S.P. Chambers and M.S.S. Su, *J. Biol. Chem.* **271** (1996), pp. 27696–27700.

[22] X. Xie, Y. Gu, T. Fox, J.T. Coll, M.A. Fleming, W. Markland, P.R. Caron, K.P. Wilson and M.S.S. Su, *Structure* **6** (1998), pp. 983–991.

[23] B.J. Canagarajah, A. Khokhlatchev, M.H. Cobb and E.J. Goldsmith, *Cell* **90** (1997), pp. 859–869.

[24] S. Bellon, M.J. Fitzgibbon, T. Fox, H.-M. Hsiao and K.P. Wilson, *Structure* **7** (1999), pp. 1057–1065.

[25] T. Zhou, M. Raman, Y. Gao, S. Earnest, Z. Chen, M. Machius, M.H. Cobb and E.J. Goldsmith, *Structure* **12** (2004), pp. 1891–1900.

[26] C.-I. Chang, B.-e. Xu, R. Akella, M.H. Cobb and E.J. Goldsmith, *Mol. Cell* **9** (2002), pp. 1241–1249.

[27] Y.S. Heo, S.K. Kim, C.I. Seo, Y.K. Kim, B.J. Sung, H.S. Lee, J.I. Lee, S.Y. Park, J.H. Kim, K.Y. Hwang, Y.L. Hyun, Y.H. Jeon, S. Ro, J.M. Cho, T.G. Lee and C.H. Yang, *EMBO J.* **23** (2004), pp. 2185–2195.

[28] A. Remenyi, M.C. Good, R.P. Bhattacharyya and W.A. Lim, *Mol. Cell* **20** (2005), pp. 951–962.

[29] R.P. Bhattacharyya, A. Remenyi, M.C. Good, C.J. Bashor, A.M. Falick and W.A. Lim, *Science* **311** (2006), pp. 822–826.

[30] S. Liu, J.-P. Sun, B. Zhou and Z.-Y. Zhang, *Proc. Natl. Acad. Sci. USA* **103** (2006), pp. 5326–5331.

[31] J.F. Ohren, H. Chen, A. Pavlovsky, C. Whitehead, E. Zhang, P. Kuffa, C. Yan, P. McConnell, C. Spessard, C. Banotai, W.T. Mueller, A. Delaney, C. Omer, J. Sebolt-Leopold, D.T. Dudley, I.K. Leung, C. Flamme, J. Warmus, M. Kaufman, S. Barrett, H. Tecle and C.A. Hasemann, *Nat. Struct. Mol. Biol.* **11** (2004), pp. 1192–1197.

[32] P.T.C. Wan, M.J. Garnett, S.M. Roe, S. Lee, D. Niculescu-Duvaz, V.M. Good, C.G. Project, C.M. Jones, C.J. Marshall and C.J. Springer, *Cell* **116** (2004), pp. 855–867.

[33] A. Khokhlatchev, S. Xu, J. English, P. Wu, E. Schaefer and M.H. Cobb, *J. Biol. Chem.* **272** (1997), pp. 11057–11062.

[34] J.L. Wilsbacher, M.H. Cobb, Methods in Enzymology, Academic Press, 332 (2001) 387–400.

[35] S.B. Patel, P.M. Cameron, B. Frantz-Wattley, E. O'Neill, J.W. Becker and G. Scapin, *Biochim. Biophys. Acta (BBA)—Proteins Proteomics* **1696** (2004), pp. 67–73.

[36] M.J. Robinson, P.C. Harkins, J. Zhang, R. Baer, J.W. Haycock, M.H. Cobb and E.J. Goldsmith, *Biochemistry* **35** (1996), pp. 5641–5646.

[37] M. Hutchison, K.S. Berman and M.H. Cobb, *J. Biol. Chem.* **273** (1998), pp. 28625–28632.

[38] J.J. Bellizzi, J. Widom, C.W. Kemp and J. Clardy, *Structure* **7** (1999), pp. R263–R267.

[39] S.J. Mansour, W.T. Matten, A.S. Hermann, J.M. Candia, S. Rong, K. Fukasawa, G.F. Vande Woude and N.G. Ahn, *Science* **265** (1994), pp. 966–970.

[40] S. Liemann, K. Chandran, T.S. Baker, M.L. Nibert and S.C. Harrison, *Cell* **108** (2002), pp. 283–295.

[41] C. Tarricone, R. Dhavan, J. Peng, L.B. Areces, L.-H. Tsai and A. Musacchio, *Mol. Cell* **8** (2001), pp. 657–669.

[42] Z.S. Derewenda, *Methods* **34** (2004), pp. 354–363.

[43] H.E. McElroy, G.W. Sisson, W.E. Schoettlin, R.M. Aust and J.E. Villafranca, *J. Crystal Growth* **122** (1992), pp. 265–272.

[44] M. Bukhtiyarova, K. Northrop, X. Chai, D. Casper, M. Karpusas and E. Springman, *Protein Expr. Purif.* **37** (2004), pp. 154–161.

[45] S.J. Kodumal, K.G. Patel, R. Reid, H.G. Menzella, M. Welch and D.V. Santi, *Proc. Natl. Acad. Sci. USA* **101** (2004), pp. 15573–15578.

[46] Z. Wang, B.J. Canagarajah, J.C. Boehm, S. Kassisa, M.H. Cobb, P.R. Young, S. Abdel-Meguid, J.L. Adams and E.J. Goldsmith, *Structure* **6** (1998), pp. 1117–1128.

[47] C.E. Fitzgerald, S.B. Patel, J.W. Becker, P.M. Cameron, D. Zaller, V.B. Pikounis, S.J. O'Keefe and G. Scapin, *Nat. Struct. Mol. Biol.* **10** (2003), pp. 764–769.

[48] G. Scapin, S.B. Patel, J. Lisnock, J.W. Becker and P.V. LoGrasso, *Chem. Biol.* **10** (2003), pp. 705–712.

[49] A.L. Gill, M. Frederickson, A. Cleasby, S.J. Woodhead, M.G. Carr, A.J. Woodhead, M.T. Walker, M.S. Congreve, L.A. Devine, D. Tisi, M. O'Reilly, L.C.A. Seavers, D.J. Davis, J. Curry, R. Anthony, A. Padova, C.W. Murray, R.A.E. Carr and H. Jhoti, *J. Med. Chem.* **48** (2005), pp. 414–426.

[50] T.L. Blundell, H. Jhoti and C. Abell, *Nat. Rev. Drug Discov.* **1** (2002), pp. 45–54.

[51] M.J. Hartshorn, C.W. Murray, A. Cleasby, M. Frederickson, I.J. Tickle and H. Jhoti, *J. Med. Chem.* **48** (2005), pp. 403–413.

[52] C. Pargellis, L. Tong, L. Churchill, P.F. Cirillo, T. Gilmore, A.G. Graham, P.M. Grob, E.R. Hickey, N. Moss, S. Pav and J. Regan, *Nat. Struct. Mol. Biol.* **9** (2002), pp. 268–272.

[53] J. Jancarik and K. Sung-Hou, *J. Appl. Crystallogr.* **24** (1991), pp. 409–411.

[54] S. Pav, D.M. White, S. Rogers, K.M. Crane, C.L. Cywin, W. Davidson, J. Hopkins, M.L. Brown, C.A. Pargellis and L. Tong, *Protein Sci.* **6** (1997), pp. 242–245.

[55] L. Tong, S. Pav, D.M. White, S. Rogers, K.M. Crane, C.L. Cywin, M.L. Brown and C.A. Pargellis, *Nat. Struct. Mol. Biol.* **4** (1997), pp. 311–316.

[56] T. Fox, J.T. Coll, X. Xie, P.J. Ford, U.A. Germann, M.D. Porter, S. Pazhanisamy, M.A. Fleming, V. Galullo, M.S.S. Su and K.P. Wilson, *Protein Sci.* **7** (1998), pp. 2249–2255.

[57] L. Shewchuk, A. Hassell, B. Wisely, W. Rocque, W. Holmes, J. Veal and L.F. Kuyper, *J. Med. Chem.* **43** (2000), pp. 133–138.

[58] A. Trejo, H. Arzeno, M. Browner, S. Chanda, S. Cheng, D.D. Comer, S.A. Dalrymple, P. Dunten, J. Lafargue, B. Lovejoy, J. Freire-Moar, J. Lim, J. McIntosh, J. Miller, E. Papp, D. Reuter, R. Roberts, F. Sanpablo, J. Saunders, K. Song, A. Villasenor, S.D. Warren, M. Welch, P. Weller, P.E. Whiteley, L. Zeng and D.M. Goldstein, *J. Med. Chem.* **46** (2003), pp. 4702–4713.

[59] J.E. Stelmach, L. Liu, S.B. Patel, J.V. Pivnichny, G. Scapin, S. Singh, C.E.C.A. Hop, Z. Wang, J.R. Strauss and P.M. Cameron, *Bioorg. Med. Chem. Lett.* **13** (2003), pp. 277–280.

[60] T. Kinoshita, M. Warizaya, M. Ohori, K. Sato, M. Neya and T. Fujii, *Bioorg. Med. Chem. Lett.* **16** (2006), pp. 55–58.

[61] N. Tamayo, L. Liao, M. Goldberg, D. Powers, Y.-Y. Tudor, V. Yu, L.M. Wong, B. Henkle, S. Middleton and R. Syed, *Bioorg. Med. Chem. Lett.* **15** (2005), pp. 2409–2413.

[62] M. Ohori, T. Kinoshita, M. Okubo, K. Sato, A. Yamazaki, H. Arakawa, S. Nishimura, N. Inamura, H. Nakajima and M. Neya, *Biochem. Biophys. Res. Commun.* **336** (2005), pp. 357–363.

[63] S.K. Laughlin, M.P. Clark, J.F. Djung, A. Golebiowski, T.A. Brugel, M. Sabat, R.G. Bookland, M.J. Laufersweiler, J.C. VanRens and J.A. Townes, *Bioorg. Med. Chem. Lett.* **15** (2005), pp. 2399–2403.

[64] A. Golebiowski, J.A. Townes, M.J. Laufersweiler, T.A. Brugel, M.P. Clark, C.M. Clark, J.F. Djung, S.K. Laughlin, M.P. Sabat and R.G. Bookland, *Bioorg. Med. Chem. Lett.* **15** (2005), pp. 2285–2289.

[65] E.L. Michelotti, K.K. Moffett, D. Nguyen, M.J. Kelly, R. Shetty, X. Chai, K. Northrop, V. Namboodiri, B. Campbell and G.A. Flynn, *Bioorg. Med. Chem. Lett.* **15** (2005), pp. 5274–5279.

[66] K.F. McClure, Y.A. Abramov, E.R. Laird, J.T. Barberia, W. Cai, T.J. Carty, S.R. Cortina, D.E. Danley, A.J. Dipesa, K.M. Donahue, M.A. Dombroski, N.C. Elliott, C.A. Gabel, S. Han, T.R. Hynes, P.K. LeMotte, M.N. Mansour, E.S. Marr, M.A. Letavic, J. Pandit, D.B. Ripin, F.J. Sweeney, D. Tan and Y. Tao, *J. Med. Chem.* **48** (2005), pp. 5728–5737.

[67] J.E. Sullivan, G.A. Holdgate, D. Campbell, D. Timms, S. Gerhardt, J. Breed, A.L. Breeze, A. Bermingham, R.A. Pauptit, R.A. Norman, K.J. Embrey, J. Read, W.S. VanScyoc and W.H.J. Ward, *Biochemistry* **44** (2005), pp. 16475–16490.

[68] M. Liu, Z. Xin, J.E. Clampit, S. Wang, R.J. Gum, D.L. Haasch, J.M. Trevillyan, C. Abad-Zapatero, E.H. Fry, H.L. Sham and G. Liu, *Bioorg. Med. Chem. Lett.* **16** (2006), pp. 2590–2594.

[69] T. Zhou, L. Sun, J. Humphreys and E.J. Goldsmith, *Structure* **14** (2006), pp. 1011–1019.

Corresponding author. Fax: +1 214 645 6387.

[1] Present address: ARIAD Pharmaceuticals, Inc., 26 Landsdowne Street

Cambridge, MA 02139-4234, USA

## BIBLIOGRAPHY

Abo, A., J. Qu, M. S. Cammarano, C. Dan, A. Fritsch, V. Baud, B. Belisle and A. Minden (1998). "PAK4, a novel effector for Cdc42Hs, is implicated in the reorganization of the actin cytoskeleton and in the formation of filopodia." Embo J 17(22): 6527-40.

Anselmo, A. N., S. Earnest, W. Chen, Y. C. Juang, S. C. Kim, Y. Zhao and M. H. Cobb (2006). "WNK1 and OSR1 regulate the Na+, K+, 2Cl- cotransporter in HeLa cells." Proc Natl Acad Sci U S A 103(29): 10883-8.

Bennett, M. J. and D. Eisenberg (2004). "The evolving role of 3D domain swapping in proteins." Structure 12(8): 1339-41.

Bennett, M. J., M. R. Sawaya and D. Eisenberg (2006). "Deposition diseases and 3D domain swapping." Structure 14(5): 811-24.

Biggins, S. (1999). "The conserved protein kinase Ipl1 regulates microtubule binding to kinetochores in budding yeast." Genes Dev. 13: 532-544.

Bishop, J. D. and J. M. Schumacher (2002). "Phosphorylation of the carboxyl terminus of inner centromere protein (INCENP) by the Aurora B kinase stimulates Aurora B kinase activity." J. Biol. Chem. 277: 27577-27580.

Bolton, M. A. (2002). "Aurora B kinase exists in a complex with survivin and INCENP and its kinase activity is stimulated by survivin binding and phosphorylation." Mol. Biol. Cell 13: 3064-3077.

Burke, D. J. and P. T. Stukenberg (2003). "The quest for the holy "G" of chromosomal passengers." Dev Cell 5(2): 187-8.

Caenepeel, S., G. Charydczak, S. Sudarsanam, T. Hunter and G. Manning (2004). "The mouse kinome: discovery and comparative genomics of all mouse protein kinases." Proc. Natl Acad. Sci. USA 101: 11707-11712.

Canagarajah, B. J., A. Khokhlatchev, M. H. Cobb and E. J. Goldsmith (1997). "Activation Mechanism of the MAP Kinase ERK2 by Dual Phosphorylation." Cell 90(5): 859-869.

Carmena, M. and W. C. Earnshaw (2003). "The cellular geography of aurora kinases." Nat Rev Mol Cell Biol 4(11): 842-54.

Cheeseman, I. M. (2002). "Phospho-regulation of kinetochore-microtubule attachments by the Aurora kinase Ipl1p." Cell 111: 163-172.

Cheetham, G. M. (2002). "Crystal structure of aurora-2, an oncogenic serine/threonine kinase." J. Biol. Chem. 277: 42419-42422.

Chen, J. (2003). "Survivin enhances Aurora-B kinase activity and localizes Aurora-B in human cells." J. Biol. Chem. 278: 486-490.

Chen, W., M. Yazicioglu and M. H. Cobb (2004). "Characterization of OSR1, a Member of the Mammalian Ste20p/Germinal Center Kinase Subfamily." J. Biol. Chem. 279(12): 11129-11136.

Chen, Y. R., C. F. Meyer, B. Ahmed, Z. Yao and T. H. Tan (1999). "Caspase-mediated cleavage and functional changes of hematopoietic progenitor kinase 1 (HPK1)." Oncogene 18(51): 7370-7.

Cohen, P. (2000). "The regulation of protein function by multisite phosphorylation [mdash] a 25 year update." Trends Biochem. Sci. 25: 596-601.

Creasy, C. L. and J. Chernoff (1995). "Cloning and characterization of a human protein kinase with homology to Ste20." J Biol Chem 270(37): 21695-700.

Dan, I., N. M. Watanabe and A. Kusumi (2001). "The Ste20 group kinases as regulators of MAP kinase cascades." Trends in Cell Biology 11(5): 220-230.

Dehouck, Y., C. Biot, D. Gilis, J. M. Kwasigroch and M. Rooman (2003). "Sequence-structure signals of 3D domain swapping in proteins." J Mol Biol 330(5): 1215-25.

Delpire, E. and K. B. Gagnon (2006). "SPAK and OSR1, key kinases involved in the regulation of chloride transport." Acta Physiol (Oxf) 187(1-2): 103-13.

Denton, J., K. Nehrke, X. Yin, R. Morrison and K. Strange (2005). "GCK-3, a newly identified Ste20 kinase, binds to and regulates the activity of a cell cycle-dependent ClC anion channel." J Gen Physiol 125(2): 113-25.

Drogen, F., S. M. O'Rourke, V. M. Stucke, M. Jaquenoud, A. M. Neiman and M. Peter (2000). "Phosphorylation of the MEKK Ste11p by the PAK-like kinase Ste20p is required for MAP kinase signaling in vivo." <u>Curr Biol</u> 10(11): 630-9.

Druker, B. J., M. Talpaz, D. J. Resta, B. Peng, E. Buchdunger, J. M. Ford, N. B. Lydon, H. Kantarjian, R. Capdeville, S. Ohno-Jones and C. L. Sawyers (2001). "Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia." <u>N Engl J Med</u> 344(14): 1031-7.

Earnshaw, W. C. and R. L. Bernat (1990). "Chromosomal passengers: towards an integrated view of mitosis." <u>Chromosoma</u> 100: 139-146.

Eckhart, W., M. A. Hutchinson and T. Hunter (1979). "An activity phosphorylating tyrosine in polyoma T antigen immunoprecipitates." <u>Cell</u> 18(4): 925-933.

Ellinger-Ziegelbauer, H., H. Karasuyama, E. Yamada, K. Tsujikawa, K. Todokoro and E. Nishida (2000). "Ste20-like kinase (SLK), a regulatory kinase for polo-like kinase (Plk) during the G2/M transition in somatic cells." <u>Genes Cells</u> 5(6): 491-8.

Farrell, P. J., R. J. Broeze and P. Lengyel (1979). "Accumulation of an mRNA and protein in interferon-treated Ehrlich ascites tumour cells." <u>Nature</u> 279(5713): 523-525.

Gupta, S., D. Campbell, B. Derijard and R. J. Davis (1995). "Transcription factor ATF2 regulation by the JNK signal transduction pathway." <u>Science</u> 267: 389-393.

Gustin, M. C., J. Albertyn, M. Alexander and K. Davenport (1998). "MAP kinase pathways in the yeast Saccharomyces cerevisiae." <u>Microbiol Mol Biol Rev</u> 62(4): 1264-300.

Hershko, A. and A. Ciechanover (1998). "THE UBIQUITIN SYSTEM." <u>Annual Review of Biochemistry</u> 67(1): 425-479.

Hicke, L. (2001). "Protein regulation by monoubiquitin." <u>Nat Rev Mol Cell Biol</u> 2(3): 195-201.

Higuchi, T. and F. Uhlmann (2003). "Cell cycle: passenger acrobatics." Nature 426(6968): 780-1.

Honda, R., R. Korner and E. A. Nigg (2003). "Exploring the functional interactions between Aurora B, INCENP, and survivin in mitosis." Mol Biol Cell 14(8): 3325-41.

Hubbard, S. R. (1997). "Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog." Embo J 16(18): 5572-81.

Hubbard, S. R., M. Mohammadi and J. Schlessinger (1998). "Autoregulatory Mechanisms in Protein-tyrosine Kinases." J. Biol. Chem. 273(20): 11987-11990.

Hubbard, S. R., L. Wei and W. A. Hendrickson (1994). "Crystal structure of the tyrosine kinase domain of the human insulin receptor." Nature 372(6508): 746-754.

Huse, M. and J. Kuriyan (2002). "The Conformational Plasticity of Protein Kinases." Cell 109(3): 275-282.

Ip, Y. T. and R. J. Davis (1998). "Signal transduction by the c-Jun N-terminal kinase (JNK) -- from inflammation to development." Current Opinion in Cell Biology 10(2): 205-219.

Jaskolski, M. (2001). "3D domain swapping, protein oligomerization, and amyloid formation." Acta Biochim Pol 48(4): 807-27.

Jentsch, S. and G. Pyrowolakis (2000). "Ubiquitin and its kin: how close are the family ties?" Trends in Cell Biology 10(8): 335-342.

Johnson, L. N., M. E. M. Noble and D. J. Owen (1996). "Active and Inactive Protein Kinases: Structural Basis for Regulation." Cell 85(2): 149-158.

Johnston, A. M., G. Naselli, L. J. Gonez, R. M. Martin, L. C. Harrison and H. J. DeAizpurua (2000). "SPAK, a STE20/SPS1-related kinase that activates the p38 pathway." Oncogene 19(37): 4290-7.

Khokhlatchev, A. V., B. Canagarajah, J. Wilsbacher, M. Robinson, M. Atkinson, E. Goldsmith and M. H. Cobb (1998). "Phosphorylation of the MAP

Kinase ERK2 Promotes Its Homodimerization and Nuclear Translocation." <u>Cell</u> 93(4): 605-615.

Kim, K. I. and D. E. Zhang (2003). "ISG15, not just another ubiquitin-like protein." <u>Biochem Biophys Res Commun</u> 307(3): 431-4.

Knight, E., Jr. and B. Cordova (1991). "IFN-induced 15-kDa protein is released from human lymphocytes and monocytes." <u>J Immunol</u> 146(7): 2280-2284.

Knighton, D. R., J. H. Zheng, L. F. Ten Eyck, V. A. Ashford, N. H. Xuong, S. S. Taylor and J. M. Sowadski (1991). "Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase." <u>Science</u> 253(5018): 407-14.

Kok, K., R. Hofstra, A. Pilz, A. van den Berg, P. Terpstra, C. Buys and B. Carritt (1993). "A Gene in the Chromosomal Region 3p21 with Greatly Reduced Expression in Lung Cancer is Similar to the Gene for Ubiquitin-Activating Enzyme." <u>Proceedings of the National Academy of Sciences</u> 90(13): 6071-6075.

Korant, B. D., D. C. Blomstrom, G. J. Jonak and E. Knight, Jr. (1984). "Interferon-induced proteins. Purification and characterization of a 15,000-dalton protein from human and bovine cells induced by interferon." <u>J. Biol. Chem.</u> 259(23): 14835-14839.

Kuramochi, S., Y. Matsuda, M. Okamoto, F. Kitamura, H. Yonekawa and H. Karasuyama (1999). "Molecular cloning of the human gene STK10 encoding lymphocyte-oriented kinase, and comparative chromosomal mapping of the human, mouse, and rat homologues." <u>Immunogenetics</u> 49(5): 369-75.

Kuramochi, S., T. Moriguchi, K. Kuida, J. Endo, K. Semba, E. Nishida and H. Karasuyama (1997). "LOK is a novel mouse STE20-like protein kinase that is expressed predominantly in lymphocytes." <u>J Biol Chem</u> 272(36): 22679-84.

Kyriakis, J. M. (1999). "Signaling by the Germinal Center Kinase Family of Protein Kinases." <u>J. Biol. Chem.</u> 274(9): 5259-5262.

Lee, K. K., M. Murakawa, E. Nishida, S. Tsubuki, S. Kawashima, K. Sakamaki and S. Yonehara (1998). "Proteolytic activation of MST/Krs, STE20-

related protein kinase, by caspase during apoptosis." <u>Oncogene</u> 16(23): 3029-37.

Levinson, A. D., H. Oppermann, L. Levintow, H. E. Varmus and J. M. Bishop (1978). "Evidence that the transforming gene of avian sarcoma virus encodes a protein kinase associated with a phosphoprotein." <u>Cell</u> 15(2): 561-572.

Liu, Y. and D. Eisenberg (2002). "3D domain swapping: As domains continue to swap." <u>Protein Sci</u> 11(6): 1285-1299.

Loeb, K. R. and A. L. Haas (1992). "The interferon-inducible 15-kDa ubiquitin homolog conjugates to intracellular proteins." <u>J. Biol. Chem.</u> 267(11): 7806-7813.

Manning, G., G. D. Plowman, T. Hunter and S. Sudarsanam (2002). "Evolution of protein kinase signaling from yeast to man." <u>Trends Biochem. Sci.</u> 27: 514-520.

Manning, G., D. B. Whyte, R. Martinez, T. Hunter and S. Sudarsanam (2002). "The protein kinase complement of the human genome." <u>Science</u> 298: 1912-1934.

McCoy, A. J., R. W. Grosse-Kunstleve, L. C. Storoni and R. J. Read (2005). "Likelihood-enhanced fast translation functions." <u>Acta Crystallogr D Biol Crystallogr</u> 61(Pt 4): 458-64.

Melzig, J., K. H. Rein, U. Schafer, H. Pfister, H. Jackle, M. Heisenberg and T. Raabe (1998). "A protein related to p21-activated kinase (PAK) that is involved in neurogenesis in the Drosophila adult central nervous system." <u>Curr Biol</u> 8(22): 1223-6.

Mohammadi, M., G. McMahon, L. Sun, C. Tang, P. Hirth, B. K. Yeh, S. R. Hubbard and J. Schlessinger (1997). "Structures of the tyrosine kinase domain of fibroblast growth factor receptor in complex with inhibitors." <u>Science</u> 276(5314): 955-60.

Neidhardt, F. C. (1987). "What the bacteriologists have learned about heat shock." <u>Genes Dev</u> 1(2): 109-10.

Page, G., D. Kogel, V. Rangnekar and K. H. Scheidtmann (1999). "Interaction partners of Dlk/ZIP kinase: co-expression of Dlk/ZIP kinase and Par-4 results in cytoplasmic retention and apoptosis." Oncogene 18(51): 7265-73.

Piechotta, K., N. Garbarini, R. England and E. Delpire (2003). "Characterization of the interaction of the stress kinase SPAK with the Na+-K+-2Cl-cotransporter in the nervous system: evidence for a scaffolding role of the kinase." J Biol Chem 278(52): 52848-56.

Piechotta, K., J. Lu and E. Delpire (2002). "Cation chloride cotransporters interact with the stress-related kinases Ste20-related proline-alanine-rich kinase (SPAK) and oxidative stress response 1 (OSR1)." J Biol Chem 277(52): 50812-9.

Pinna, L. A. and M. Ruzzene (1996). "How do protein kinases recognize their substrates?" Biochim. Biophys. Acta 1314: 191-225.

Pru, J. K., K. J. Austin, A. L. Haas and T. R. Hansen (2001). "Pregnancy and Interferon-{tau} Upregulate Gene Expression of Members of the 1-8 Family in the Bovine Uterus." Biol Reprod 65(5): 1471-1480.

Ptacek, J. (2005). "Global analysis of protein phosphorylation in yeast." Nature 438: 679-684.

Read, R. J. (2001). "Pushing the boundaries of molecular replacement with maximum likelihood." Acta Crystallogr D Biol Crystallogr 57(Pt 10): 1373-82.

Recht, M., E. C. Borden and E. Knight, Jr. (1991). "A human 15-kDa IFN-induced protein induces the secretion of IFN-gamma." J Immunol 147(8): 2617-2623.

Rousseau, F., J. W. Schymkowitz and L. S. Itzhaki (2003). "The unfolding story of three-dimensional domain swapping." Structure 11(3): 243-51.

Sabourin, L. A. and M. A. Rudnicki (1999). "Induction of apoptosis by SLK, a Ste20-related kinase." Oncogene 18(52): 7566-75.

Sabourin, L. A., P. Seale, J. Wagner and M. A. Rudnicki (2000). "Caspase 3 Cleavage of the Ste20-Related Kinase SLK Releases and Activates an

Apoptosis-Inducing Kinase Domain and an Actin-Disassembling Region." Mol. Cell. Biol. 20(2): 684-696.

Scheeff, E. D. and P. E. Bourne (2005). "Structural evolution of the protein kinase-like superfamily." PLoS Comput Biol 1(5): e49.

Scheel, H. and K. Hofmann (2003). "A novel interaction motif, SARAH, connects three classes of tumor suppressor." Curr Biol 13(23): R899-900.

Schindler, T., W. Bornmann, P. Pellicena, W. T. Miller, B. Clarkson and J. Kuriyan (2000). "Structural mechanism for STI-571 inhibition of abelson tyrosine kinase." Science 289(5486): 1938-42.

Sells, M. A. and J. Chernoff (1997). "Emerging from the Pak: the p21-activated protein kinase family." Trends in Cell Biology 7(4): 162-167.

Shani, G., L. Marash, D. Gozuacik, S. Bialik, L. Teitelbaum, G. Shohat and A. Kimchi (2004). "Death-associated protein kinase phosphorylates ZIP kinase, forming a unique kinase hierarchy to activate its cell death functions." Mol Cell Biol 24(19): 8611-26.

Storbeck, C. J., K. Daniel, Y. H. Zhang, J. Lunde, A. Scime, A. Asakura, B. Jasmin, R. G. Korneluk and L. A. Sabourin (2004). "Ste20-like kinase SLK displays myofiber type specificity and is involved in C2C12 myoblast differentiation." Muscle Nerve 29(4): 553-64.

Storoni, L. C., A. J. McCoy and R. J. Read (2004). "Likelihood-enhanced fast rotation functions." Acta Crystallographica Section D 60(3): 432-438.

Strange, K., J. Denton and K. Nehrke (2006). "Ste20-type kinases: evolutionarily conserved regulators of ion transport and cell volume." Physiology (Bethesda) 21: 61-8.

Taylor, G. (2003). "The phase problem." Acta Crystallogr D Biol Crystallogr 59(Pt 11): 1881-90.

Taylor, L. K., H. C. Wang and R. L. Erikson (1996). "Newly identified stress-responsive protein kinases, Krs-1 and Krs-2." Proc Natl Acad Sci U S A 93(19): 10099-104.

Thomas, J. G. and F. Baneyx (1997). "Divergent effects of chaperone overexpression and ethanol supplementation on inclusion body formation in recombinant Escherichia coli." <u>Protein Expr Purif</u> 11(3): 289-96.

Ubersax, J. A. (2003). "Targets of the cyclin-dependent kinase Cdk1." <u>Nature</u> 425: 859-864.

VanBogelen, R. A., P. M. Kelley and F. C. Neidhardt (1987). "Differential induction of heat shock, SOS, and oxidation stress regulons and accumulation of nucleotides in Escherichia coli." <u>J Bacteriol</u> 169(1): 26-32.

Vitari, A. C., J. Thastrup, F. H. Rafiqi, M. Deak, N. A. Morrice, H. K. Karlsson and D. R. Alessi (2006). "Functional interactions of the SPAK/OSR1 kinases with their upstream activator WNK1 and downstream substrate NKCC1." <u>Biochem J</u> 397(1): 223-31.

Wagner, S., T. A. Flood, P. O'Reilly, K. Hume and L. A. Sabourin (2002). "Association of the Ste20-like kinase (SLK) with the microtubule. Role in Rac1-mediated regulation of actin dynamics during cell adhesion and spreading." <u>J Biol Chem</u> 277(40): 37685-92.

Walter, S. A., R. E. Cutler, Jr., R. Martinez, M. Gishizky and R. J. Hill (2003). "Stk10, a new member of the polo-like kinase kinase family highly expressed in hematopoietic tissue." <u>J Biol Chem</u> 278(20): 18221-8.

Widmann, C., S. Gibson, M. B. Jarpe and G. L. Johnson (1999). "Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human." <u>Physiol Rev</u> 79(1): 143-80.

Willoughby, E. A., G. R. Perkins, M. K. Collins and A. J. Whitmarsh (2003). "The JNK-interacting protein-1 scaffold protein targets MAPK phosphatase-7 to dephosphorylate JNK." <u>J. Biol. Chem.</u> 278: 10731-10736.

Wu, C., M. Whiteway, D. Y. Thomas and E. Leberer (1995). "Molecular characterization of Ste20p, a potential mitogen-activated protein or extracellular signal-regulated kinase kinase (MEK) kinase kinase from Saccharomyces cerevisiae." <u>J Biol Chem</u> 270(27): 15984-92.

Wu, S., J. Huang, J. Dong and D. Pan (2003). "hippo encodes a Ste-20 family protein kinase that restricts cell proliferation and promotes apoptosis in conjunction with salvador and warts." Cell 114(4): 445-56.

Yamaguchi, H. and W. A. Hendrickson (1996). "Structural basis for activation of human lymphocyte kinase Lck upon tyrosine phosphorylation." Nature 384(6608): 484-489.

Yamaguchi, H., M. Matsushita, A. C. Nairn and J. Kuriyan (2001). "Crystal structure of the atypical protein kinase domain of a TRP channel with phosphotransferase activity." Mol Cell 7(5): 1047-57.

Yergeau, D. A., C. J. Hetherington, Q. Wang, P. Zhang, A. H. Sharpe, M. Binder, M. Marin-Padilla, D. G. Tenen, N. A. Speck and D.-E. Zhang (1997). "Embryonic lethality and impairment of haematopoiesis in mice heterozygous for an AML1-ETO fusion gene." Nat Genet 15(3): 303-306.

Zhang, F., A. Strand, D. Robbins, M. H. Cobb and E. J. Goldsmith (1994). "Atomic structure of the MAP kinase ERK2 at 2.3 A resolution." Nature 367: 704-711.

Zhao, B., M. J. Bower, P. J. McDevitt, H. Zhao, S. T. Davis, K. O. Johanson, S. M. Green, N. O. Concha and B.-B. S. Zhou (2002). "Structural Basis for Chk1 Inhibition by UCN-01." J. Biol. Chem. 277(48): 46609-46615.

Zhou, T., M. Raman, Y. Gao, S. Earnest, Z. Chen, M. Machius, M. H. Cobb and E. J. Goldsmith (2004). "Crystal Structure of the TAO2 Kinase Domain: Activation and Specificity of a Ste20p MAP3K." Structure 12(10): 1891-1900.

Zhu, H. (2000). "Analysis of yeast protein kinases using protein chips." Nature Genet. 26: 283-289.

**VITAE**


Seung-Jae Lee was born in Seoul, Korea, on Sep 16, 1973.  He is the son of Jun-Seop Lee and Suk-Kyung Sung, and the brother of Seung-Ah Lee.  After completing his work at Dae-Won Foreign Language High School in 1992, he entered Han-Kuk University of Foreign Studies in Seoul, Korea, majoring in French education.  Following the completion of military service in Korea over 26 months in 1996, he transferred to Iowa State University, majoring in biochemistry.  After completing the degree of Bachelor of Science with a major in biochemistry (graduation with distinction) from Iowa State University, he entered the graduate school of biomedical sciences at the University of Texas Southwestern Medical Center at Dallas in 2001.  He was admitted to Ph.D candidacy in the Molecular Biophysics program in 2003.  In 1999, he married Soon-Hee Sul.  Their son, Joshua A. Lee was born in 2004.



Permanent Address:     3633 W.Walnut Hill Ln Apt #1092

                       Irving, TX 75038