Stony Brook University

Lack of Csk-mediated negative regulation in a unicellular Src kinase

A Dissertation Presented

by

Kira Schultheiss

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Molecular and Cellular Biology

Stony Brook University

August 2013

Stony Brook University

The Graduate School

Kira Schultheiss

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

Todd Miller, Ph.D., Dissertation Advisor
Professor, Department of Physiology and Biophysics

Mark Bowen, Ph.D., Chairperson of Defense
Assistant Professor, Department of Physiology and Biophysics

Nancy Reich Marshall, Ph.D.
Professor, Department of Molecular Genetics and Microbiology

Elizabeth Boon, Ph.D.
Assistant Professor, Department of Chemistry

Gerald Thomsen, Ph.D.,
Professor, Department of Biochemistry and Cell Biology

Raafat El-Maghrabi, Ph.D.
Research Associate Professor, Department of Physiology and Biophysics,
Stony Brook University

This dissertation is accepted by the Graduate School

Charles Taber
Interim Dean of the Graduate School

Abstract of the Dissertation

Lack of Csk-mediated negative regulation in a unicellular Src kinase.

by

Kira Schultheiss

Doctor of Philosophy

in

Molecular and Cellular Biology

Stony Brook University

2013

Tyrosine kinases play important roles in cellular communication in animals. c-Src is a nonreceptor tyrosine kinase that is involved in survival, angiogenesis, proliferation, and invasion pathways. There are 9 Src family kinases, and they have a conserved domain arrangement (SH3-SH2-kinase) that is necessary for regulation. The basal activity of c-Src is normally kept low due to phosphorylation by C-terminal Src kinase (Csk). This regulatory relationship has been conserved throughout the evolution of animals. Until recently, it was thought that complex signaling molecules, such as the tyrosine kinases, did not exist outside of multicellular animals. Newly discovered unicellular precursors to mammalian signaling proteins are hypothesized to have played a major role in the transition from unicellular organisms to multicellularity.

Homologs of tyrosine kinases have been found in *Capsaspora owczarzaki* and *Ministeria vibrans*, two filasterean unicellular organisms. The genomes of *C. owczarzaki* and *M. vibrans* each contain two Src-like kinases and one Csk kinase. We have cloned, expressed and purified all three proteins in *C. owczarzaki* (CoSrc1, CoSrc2 and CoCsk) and *M. vibrans* (MvSrc1, MvSrc2 and MvCsk) and examined their properties. CoSrc1, CoSrc2, MvSrc1 and MvSrc2 are all active towards synthetic tyrosine kinase substrates, and display different levels of substrate targeting and specificity. Surprisingly, both CoCsk and MvCsk lack activity toward a general

tyrosine kinase substrate. Neither Csk phosphorylates or inhibits its respective Src kinase. CoSrc kinases in *Capsaspora* cells are active, consistent with an absence of negative regulation. The tight regulation of Src was an integral step in the development of proper cell-cell interaction in multicellular organisms, and appears to have developed later at the onset of metazoan evolution.

We have also carried out preliminary studies aimed at identifying Src substrates in choanoflagellates. One substrate in *M. brevicollis* is the C-terminus of a receptor tyrosine kinase, RTKB2. This RTK contains 6 copies of a unique motif that is phosphorylated by the SFK MbSrc1. Additionally, we studied a homologue of STAT, a transcription factor that is activated by Src in metazoans, from the choanoflagellate *Salpingocea rosetta*. Characterizing more, diverse signaling molecules from unicellular organisms will provide insights into one of the most important events in the evolution of Metazoa: the transition to multicellularity.

Table of Contents

List of Figures

.

List of Abbreviations

ATP	Adenosine triphosphate
cDNA	Complementary DNA
CNBr	Cyanogen bromide
CoCsk	*Capsaspora owczarzaki* Csk
CoSrc1/2	*Capsaspora owczarzaki* Src1/2
Csk	C-terminal Src kinase
DMEM	Dulbecco's modified eagle's medium
EDTA	Ethylene diamine tetra acetate
EGFR	Epidermal growth factor receptor
FBS	Fetal bovine serum
GFP	Green fluorescent protein
GST	Glutathione S-transferase
HRP	Horse radish peroxidase
IP	Immunoprecipitation
IPTG	Isopropyl β-D-1-thiogalactopyranoside
MbCsk	*Monosiga brevicollis* Csk
MbSrc1/4	*Monosiga brevicollis* Src1/2
Mg/mL	milligram per milliliter
$MgCl_2$	Magnesium chloride
mL	Milliliter
mM	Millimolar
MS	Mass spectrometry
MvCsk	*Ministeria vibrans* Csk
MvSrc1/2	*Ministeria vibrans* Src1/2
nM	Nanomolar
NaCl	Sodium chloride
nM	Nanomolar
NP40	Nonidet P40
NRTK	Nonreceptor tyrosine kinase
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PTB	Phosphotyrosine-binding
pTyr	Phospho-tyrosine
pY	Phospho-tyrosine
PVDF	Polyvinylene difluoride
RTK	Receptor tyrosine kinase
S.D.	Standard deviation
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Sf9	*Spodoptera frugiperda* 9
SFK	Src family kinase
SH2	Src homology domain 2
SH3	Src homology domain 3
SYF	Src, Yes, Fyn knockout cell

TK	Tyrosine kinase
µg	Micrograms
µL	Microliter
µM	Micromolar
YOP	*Yersinia* tyrosine phosphatase

Acknowledgments

I would like to thank my advisor, Dr. Todd Miller for giving me the opportunity to be a part of his lab. I learned a great deal from his mentorship, not only in science, but in time-efficiency and organizational skills. He has been a great example for how to have a successful career and love what you do, but also to balance work with home and have a well-rounded life. He is always open and available to students, and fosters a real sense of camaraderie in the lab with traditions of birthday cakes and lab parties at his house. I could not have asked for a better mentor as I went through the stressful graduate school years, and can't imagine having done it in any other lab.

I want to thank my committee members for their discussions and support in committee meetings. Serving as my committee chair, Dr. Mark Bowen has been invaluable with suggestions and advice from the proposal all the way up to the thesis. It is nice to have such a friendly, collaborative lab right next door to my own. Dr. Elizabeth Boon and Dr. Gerald Thomson have been very insightful with comments and feedback, and always go out of their way to adjust their schedules for my meetings. I thank Dr. Nancy Reich for all of her feedback, as well as her generous offers of help whenever I am having trouble with luciferase assays, or need a vector. I would like to thank Dr. Raafat El-Maghrabi for his kindness and agreeing to review my thesis, as well as sharing equipment when our thermocycler is not working.

I am very grateful to our collaborators, Dr. Hiroshi Suga and Dr. Inaki Ruiz-Trillo from the University of Barcelona. They were wonderful with the sharing of *Capsaspora* cells and information, and Dr. Suga in particular always took the time to answer all of my questions about these unfamiliar organisms, and give me feedback on posters for seminars. Although I never met him in person, I appreciate the kindnesses he showed me over email. I also would like to thank Tera Levin and Dr. Nicole King from the University of California at Berkeley for sharing both synthetic peptides and information about *S. rosetta* STAT. I would also like to thank Dr. Markus Seeliger and Michael Tong from Stony Brook University for their collaboration.

I am so thankful that I was able to work with everyone in the Miller Lab. The scientific productivity was wonderful, but what I will always remember is the warm atmosphere that exists, making it a pleasure to come to work every day. Barbara Craddock plays a large part in setting the tone for such a friendly, welcoming lab, and I will never be able to properly thank her for all the help she has given me over the years. Victoria and Wanqing were the senior students when I joined, and I will always appreciate how they showed me the ropes, and discussed life with me. I want to thank Tiffany for her friendship and mutual love of reality television, Saadat for his political debates, Richie for his endless patience for computer assistance and experiment ideas, and Zulema for her constant advice and career-building assistance. Every member of the lab has been nothing but kind and helpful, and I have appreciated that every day that I came to work.

Finally, I would like to thank my family and friends that have supported me over the years. My parents have encouraged me at an early age to work hard and aspire to the best of my abilities. They instilled the value of education into me since I was a young child, and they are the main reason I am here today. They are my biggest supporters, academically as well as personally, and I could never begin to explain how important they are to me. I would also like to thank my brother Adam and sister-in-law Colleen, who have always encouraged and assisted me in every way possible, and who I count as two of my best friends. I also would like to acknowledge my best friend Lisa, who not only has supported me and cheered me on throughout graduate school, but has been there for me since middle school. Finally, I would like to thank my boyfriend Ross, who is always willing to do anything it takes to support me and allow me to succeed in my endeavors. I cannot thank him enough for all of the support and patience he has shown throughout the writing of this thesis.

Chapter 1

Introduction

Introduction

Protein Kinases

A kinase is an enzyme that transfers phosphate groups from high-energy donor molecules (ATP) to specific substrates. Protein phosphorylation is a critical part of signal transduction in prokaryotes and eukaryotes. In prokaryotes, phospho-histidine signaling predominates, while serine, threonine, and tyrosine signaling is more common in eukaryotic organisms [1, 2]. Phosphotyrosine signaling appears to have emerged more recently in the evolutionary tree. In addition to their absence from prokaryotes, eukaryotic tyrosine kinases are not seen in plants, yeast, fungi, and the slime mold *Dictyostelium discoideum* [3]. Tyrosine kinases are vital in the regulation of growth, differentiation, motility, survival and cell-cell communication in multicellular organisms [1, 4].

The tyrosine kinases (TKs) are a large, well-studied, multigene family of the protein kinases (Figure 1-1). There are 90 TK genes and 5 pseudogenes in humans. Of these, fifty-eight are receptor tyrosine kinases (RTKs), which are further divided into 20 subfamilies based on sequence homology. RTKs are characterized by ligand-binding domains in the extracellular region, a single transmembrane region, and an intracellular region containing the kinase domain as well as other regulatory domains[5]. The extracellular domains on the cell surface receive environmental information through high-affinity ligand binding. This causes a conformational change that releases autoregulation, activating the intracellular kinase domain. This leads to autophosphorylation, which promotes further conformational changes. These changes are important for enzymatic activity, as well as recruitment of other cytoplasmic proteins containing Src Homology 2 (SH2) domains such as Src, RasGAP, SHP-1 and PI(3)K [6, 7]. RTKs respond to a variety of extracellular stimuli, initiating signals that regulate important cellular and developmental processes [8, 9]. More than half (32) of the known RTKs have been observed in mutated or overexpressed forms, leading to human disease. Point mutations in RTKs can cause an inability to properly form an inactive conformation, leading to increased activity in the absence of ligand stimulation. The overexpression of the RTKs is also linked to constitutively activated kinases by increasing the numbers of dimers. For example, ErbB2 and epidermal growth factor receptor (EGFR) are often overexpressed in breast and lung cancers [8].

The remaining 32 TK genes encode non-receptor tyrosine kinases (nRTKs), which have 10 sub-families. The nRTKs are integral parts of the signaling cascade, triggered by RTKs or other receptors such as G protein-coupled receptors and receptors from the immune system. The nRTKs include enzymes such as the Janus kinases (Jaks), Abl, and the largest family of nRTKs, the Src kinases (Figure 1-1) [3, 10].

Src Family Kinases (SFKs)

The SFKs are a family of nonreceptor tyrosine kinases that act as key mediators of cellular signaling. There are 9 members: Src, Yes, Fyn, Fgr, Lck, Hck, Blk, Lyn and Yrk, which are involved in cellular proliferation, survival, migration, and growth factor and cytokine stimulation pathways [11]. Most SFKs are expressed mainly in hematopoietic cells, though Src, Yes and Fyn have a much more ubiquitous expression pattern. The SFKs are absent from bacteria, plants, or fungi, and are thought to play integral roles in cell-cell interaction in multicellular animals. A homologue of the cellular Src, v-Src, was the first proto-oncogene to be discovered. Among other mutations, v-Src contains a C-terminal sequence divergence which disrupts c-Src regulation, increasing the basal level of activity [12]. Aberrant activity of the SFKs has been linked to several human malignancies, including breast, lung and colon cancers. For instance, c-Src protein levels and kinase activity are elevated in colon carcinoma (5-8 fold higher activity) as compared to normal samples, and c-Src activation contributes to the progression of colon cancer towards the metastatic phenotype [13]. SFKs are able to interact with receptor tyrosine kinases (RTKs), both activating and being activated by the RTKs [14].

The SFKs have a characteristic domain arrangement with a myristoylation sequence at the N-terminus, followed by a unique region, SH3 and SH2 domains, a linker region, a highly conserved kinase domain, and a short C-terminal regulatory region with an important tyrosine (Y527 in Src) (Figure 1-2).

Myristoylation is critical for association of Src kinases with cellular membranes [15]. This targeting localizes Src to the correct area of the cell in order to interact with its binding partners and substrates [16]. Myristoylation has also been shown to increase c-Src kinase activity, as well as regulate its ubiquitination and degradation [17].

The unique region is 50-70 residues, and is divergent among family members. This region has varying roles in different SFK members, some of them still unclear. For example, in Lck, the unique domain interacts with the cytoplasmic tails of T-cell receptors CD4 and CD8 [18, 19]. In Fyn, this region interacts with the cytoplasmic chains of the T-cell receptor through its amino-terminal region [20]. The unique domains, though not always fully understood, are necessary for proper SFK functioning. Chimeric studies of Src and Yes show that the Src unique domain cannot be substituted by the corresponding Yes domain, indicating specific interactions and functions of this domain in each family member [21].

The SH3 domain is comprised of approximately 60 amino acids, and is easily identified through sequence similarity. More than 300 proteins in the human kinome contain SH3 domains, from cytoskeletal components to signaling molecules. The SH3 domain was first recognized as regions of sequence similarity in divergent proteins such as the SFK members, Crk adaptor protein, and phospholipase C-γ. All SH3 domains contain two packed β-sheets at right angles to each other. There are two variable loops on the ligand-binding side (the RT and N-Src), and one loop on the opposite side, referred to as the distal loop. The SH3 domain has a patch of hydrophobic residues that makes it a good binding partner of ligand sequences with a PxxP motif. This motif will bind to the SH3 domain, adopting a polyproline type II helical conformation [22]. An illustrative example of SH3 functioning is its role in the nucleation of new actin filaments at a specific time and place in the cell. A large conglomerate of proteins such as WASP, Nck, Myosin 1 and Pak are able to form a complex due to numerous SH3-mediated and other protein-protein interactions. The ubiquitous nature of the SH3 domains, as well as their relatively low affinity and specificity, allow them to promote large protein complexes, which can also disassemble quickly upon changes in environment [23].

The SH2 domain is a module of approximately 100 amino acids that binds to specific phosphotyrosine (pY) containing motifs. The SH2 domain is found within 115 human proteins, allowing domain-containing proteins to localize to areas of tyrosine phosphorylation. pY motifs are prevalent in signaling proteins, as addition of a phosphate to tyrosine is the main mechanism for tyrosine kinases to regulate intracellular processes. SH2 domains have a central anti-parallel β-sheet that is flanked by two α-helices [24]. The domain has a highly conserved pocket that will recognize the phosphorylated tyrosine of a substrate, and a more varied pocket that will

interact with the 3-6 amino acids C-terminal to the pY, thus imposing substrate specificity among the different SH2-containing proteins. The SH2 domain is classified into two main groups: group I and group III. Group I proteins, such as Crk and Nck, prefer phosphopeptide sequences with an acidic/hydrophilic residue at the pY+1 residue and a hydrophobic residue at pY+3. Group III SH2 domains, such as seen in PI3 kinase and PLC-γ, prefer hydrophobic residues at both positions. When the phosphotyrosine is recognized, it binds as an extended β-strand at a 90° angle to the SH2 β-sheet. A ligand with optimal sequence will bind to the SH2 domain with a dissociation constant of 200-600 nM [25].

The kinase domain of SFKs is highly conserved between family members, as well as other tyrosine kinases [4]. The N-terminal (small) lobe has five β-strands and one α-helix, the C helix. The C-terminal (large) lobe is mostly α-helical, and contains the activation loop. This activation loop has a critical tyrosine residue (Y416), which when phosphorylated will increase the kinase activity of the signaling protein. Nucleotide binding and phosphotransfer occur in a cleft between the two kinase domain lobes [10, 26].

The C-terminal tail of all of the SFKs has a conserved tyrosine (Y527) that is important for protein regulation. When Src is phosphorylated on this residue by C-terminal Src kinase (Csk), kinase activity decreases. The SH2 domain has an affinity for this pY, forming an intramolecular interaction that down-regulates the protein.

SFK Substrate Specificity

Studies with synthetic peptides [27] and synthetic peptide libraries [28] have indicated preferences for specific amino acids surrounding the phosphorylated tyrosine. However, the intrinsic specificity of tyrosine kinase catalytic domains does not appear to be high enough to explain the selective signaling observed in cells. The SH3 and SH2 domains are essential in determining which substrates will be recognized and phosphorylated by SFKs. Substrate peptides that contain SH2 or SH3 ligands are better phosphorylated by SFKs than control peptides that lack these recognition sequences [29]. This is further supported by protein engineering studies in which novel SFKs were created by replacing the SH2 or SH3 domains of Hck with a PDZ domain. The PDZ-kinases phosphorylated different cellular substrates, showing that an associated domain can control the substrate specificity of a kinase domain [30].

Additionally, when the SH2 domain of Abl was replaced with corresponding domains from Ras-GTPase-activating protein, Src, or Crk, the new heterologous domains are able to compensate from the missing Abl SH2, but they induced altered patterns of tyrosine-phosphorylated proteins *in vivo* [31]. Most known SFK substrates have either SH2 or SH3 ligands, with the majority containing both. Some commonly known substrates of SFKs include Cas, Sin, AFAP110, and Sam68, all of which exhibit both SH2 and SH3 ligands [4, 32-34] .

The domain arrangement of SFKs has been highly conserved throughout evolution. In addition to playing an important role in autoregulation, the proper order of the SFK domains affects substrate specificity. When Src mutants were produced with varied domain order, it led to specific changes in substrate specificity as well as phosphorylation efficiency [35].

SFK Regulation

The non-catalytic domains of SFKs, in addition to contributing to substrate specificity, are integral in the regulation of kinase activity. Src and Hck crystal structures have been determined with constructs containing the SH3, SH2, and kinase domains, with the C-terminal tail intact. The myristoylation sequence and unique domains were not included in the crystal structures, but the SH3-SH2 domains appear to have all the requirements necessary for the SFK to be in an autoinhibited state. There are two intramolecular interactions that are sufficient to inhibit the kinase: a binding of the SH2 domain to the phosphorylated C-terminal tail, and an interaction between the SH3 domain and a polyproline helix in the SH2-kinase linker region (Figure 1-2) [36, 37]. These interactions cause conformational changes in the activation loop which reduce the kinase activity. The first change is to close the ATP-binding cleft at the subdomain interface of the kinase domain, as compared to a more open conformation in an active kinase [38]. There is also an outward rotation of an αC-helix in the N-terminal lobe of the kinase domain which disrupts formation of an important salt bridge when the protein is autoinhibited. Lastly, the actual position of the activation loop shifts, blocking the substrate binding pocket. This prevents complete formation of the peptide recognition site and buries the critical activation loop tyrosine between the two kinase lobes, keeping it unphosphorylated. The orientations of the autoinhibited kinase lobes vary slightly in different SFK members, and it does not seem that complete exclusion of the nucleotide occurs upon down-regulation (as opposed to other kinases,

such as insulin receptor kinase). Rather, the affinities of autoinhibited kinases change towards ATP and ADP, shifting the equilibrium towards low kinase activity [39, 40].

Activation of the SFKs occurs through three main pathways, which disrupt the intramolecular autoinhibitory reactions discussed above (Figure 1-3). SFK activation can occur through SH3 displacement, SH2 displacement, or dephosphorylation of the C-terminal tail tyrosine [41]. Numerous proteins are activated by ligand-induced SH3 displacement. One example is the activation of Hck by Nef, which contains a high-affinity ligand for Hck's SH3 domain. Nef is an HIV accessory protein that is necessary for AIDS progression [42]. While Nef promotes autophosphorylation of Hck in its activation loop, a mutant Nef lacking a polyproline sequence elicits no activation of the kinase [43]. This experiment highlights the specific role of the SH3 domain ligand in releasing SFK autoinhibition.

The SFKs can act as activators by phosphorylating other nonreceptor tyrosine kinases on their SH3 domains. The Abl kinase inhibitor imatinib is very effective in inhibiting the activity of Bcr-Abl, the tyrosine kinase oncogene responsible for chronic myelogenous leukemia (CML). However, clinical resistance to imatinib occurs not only as a result of domain mutations in Abl, but also from over-activation of the Src family kinases in CML cells [44, 45]. The SFKs phosphorylate the Bcr-Abl SH3 domain, interfering with docking of the SH3 onto the SH2-kinase linker region, and increasing oncogenic activity. When Abl kinase variants were created with a stronger SH3:linker interaction, activity decreased greatly, and sensitivity to inhibitor increased. Screening assays for small molecules that could therapeutically induce the same effect are underway [46, 47].

SH2 ligation activates SFKs in a similar fashion to SH3 ligation (Figure 1-3). Functional evidence of the activation of c-Src through its SH2 domain is shown through studies of the epidermal growth factor receptor (EGFR) and platelet-derived growth factor receptor (PDGFR). When EGF is added to cells overexpressing EGFR, there is rapid and persistent activation of c-Src. EGFR associates with Src through its SH2 domain, releasing Src's autoinhibition and increasing its kinase activity [4, 14]. Once activated, Src can phosphorylate STAT proteins, thus propagating a signaling pathway that can start at the cell surface and continue all the way to the nucleus [48].

The intramolecular interactions that keep Src kinases down-regulated have a relatively low affinity. For example, the SH2 domain ligand in the C-terminal tail is not an optimal sequence for high-affinity binding. This low-affinity binding is necessary for the ability of outside ligands to compete, thus activating the kinase. A mutant Hck was engineered to contain the high-affinity pYEEI SH2 ligand sequence in the C-terminal tail. This mutant could not be activated by SH2 ligands, highlighting the importance of the proper balance of affinities. The SFKs need the ability to be autoinhibited, yet the down-regulation must be readily reversible when kinase activity is needed [49].

The third way Src kinases can be activated is through dephosphorylation of the C-terminal tail kinase by a phosphatase. SFK C-terminal specific phosphatases that recognize the inhibitory carboxy-terminal tyrosine will cause dephosphorylation of the site. This dephosphorylation destabilizes the down-regulated form of the protein, increasing kinase activity [4, 11]. For instance, CD45, a leukocyte-specific transmembrane protein, will activate Src kinases associated with B-cell and T-cell antigen receptor signaling by constitutive dephosphorylation of the tyrosine [50].

Csk

Csk was identified in the late 1980's as the enzyme that specifically phosphorylates Src family members at their C-terminal tail tyrosine, helping to establish their low basal activity [51]. The regulatory relationship of Src and Csk has been highly conserved throughout the evolution of animals [52, 53]. The functional importance of Csk in maintaining SFKs at a basal level in normal cells is highlighted in mutant studies. Mice with targeted disruption of the Csk gene died *in utero* and exhibited defects in neural development [54, 55]. Conversely, when both Csk and Src are knocked out, development is more successful, and near-normal neural tubes are seen [56].

Csk lacks a myristoylation sequence but contains SH3, SH2 and kinase domains in a similar arrangement as Src kinases. On the other hand, the X-ray structure shows a very different pattern of intramolecular interactions than those involved in Src auto-regulation (Figure 1-4) [12, 57]. Src kinase domain is highly active in the absence of its regulatory domains. Conversely the Csk kinase domain has very low activity when isolated, and regains some of its

activity when mixed with SH3-SH2 constructs [58-60]. This is because the active form of Csk requires specific interactions between the SH2/SH3 domains and the kinase domain. Unlike Src kinases, the Csk SH2 and SH3 domains dock onto the kinase domain N-terminal lobe, with the SH2-SH3 linker region interacting with α-helix C of the kinase domain. The displacement of the α-helix C determines whether Csk is in an active or inactive form [61].

Although Csk has the same basic domain structure as SFKs, there are some key differences. Csk lacks an N-terminal membrane-targeting region, as well as the critical tyrosines in the activation loop and the C-terminal tail that are important in Src activity and regulation. Since Csk does not contain a myristoylation sequence, it is predominantly a cytosolic protein. Its substrate, the Src kinases, all contain membrane-anchoring sequences, and therefore Csk must be transported to the membrane in order to impose down-regulation. This is done by interacting with membrane or signaling adaptor proteins through the SH2 or SH3 domains. For instance, Csk can be translocated to the membrane by Csk-binding protein (Cbp), caveolin-1, or CagA, all through interaction with the SH2 domain of Csk [62-64]. In all three cases, activated Src kinases phosphorylate these intermediary proteins, activating a negative regulatory cascade. The phosphorylated kinases bind to Csk's SH2 domain, transporting it to the membrane, where Csk can phosphorylate the Src kinases, de-activating them. In addition, Csk interacts with phosphatases through its SH3 domain to reinforce its negative regulation of SFKs. In T cells, Csk associates with the tyrosine phosphatase PEP. PEP dephosphorylates the activation loop tyrosine of the Src family kinase Lck at the same time that Csk imposes its negative regulatory phosphorylation of Lck's C-terminal tail [65].

Unlike most protein kinases, which phosphorylate multiple substrates *in vivo*, Csk only has one specific substrate: the C-terminal tyrosine of SFKs (e.g. Y527) [51, 66]. This strict specificity has been widely researched. Biochemical studies have shown that while Csk efficiently phosphorylates intact Src kinases, its activity towards synthetic peptides derived from the C-terminal tail sequences is 10,000-fold less efficient [67, 68]. The specificity of Csk for Src kinases can be explained in two main ways, and highlights the close evolution of the two proteins. The first reason is that Csk has a deletion in its activation loop which prevents it from phosphorylating most proteins: the shortened loop makes it difficult for a protein to interact with residues important for catalysis and peptide substrate binding. The second reason Csk is specific

for Src is that the SFKs have evolved so that a protein-protein interaction between Src and Csk on the C-terminal lobes of the kinase domains inserts the Src C-terminal tail directly into the activation loop. By orienting the tail at the direct angle into the activation loop, Src is able to be phosphorylated at an efficiency that other proteins cannot [69].

Pseudokinases

The term "pseudokinase" arose to describe kinases that are predicted to lack certain critical residues, rendering them enzymatically inactive. Approximately 10% of proteins in the human kinome were determined to be pseudokinases, based on the absence of three critical motifs in the catalytic domain: VAIK, and HRD, and DFG for ATP and peptide binding [70]. Although previously thought to serve solely as inactive scaffolding proteins, new studies have found activity in some of the pseudokinases. For instance, the WNK kinase is classified as a pseudokinase, but has been shown to be enzymatically active through an alternative mechanism. Although WNK lacks the critical lysine (K) in VAIK, a neighboring lysine from the catalytic subdomain II can be substituted [71, 72]. Likewise, Ca^{2+}/calmodulin-dependent serine protein kinase (CASK) does not contain the DFG motif necessary for Mg^{2+} binding. Even so, it is able to adapt an active conformation, and both autophosphorylate and phosphorylate specific substrates in the absence of Mg^{2+} [73, 74]. Whatever role pseudokinases play, either as inactive scaffolds, activators of other proteins, or active kinases, evidence has been shown that they are important in cell proliferation, differentiation, apoptosis and metastasis [75]. The varied placement of pseudokinases throughout the evolutionary tree indicates that they have similar roles as protein kinases, and that they may have been evolutionary counterparts of protein kinases.

Evolution of Multicellular Animals

Approximately 600 million years ago, the Cambrian explosion brought the emergence of multicellular animals [76-78]. This multicellularity necessitated a system for cells to communicate between themselves within an organism. Signaling proteins such as TKs were necessary for this transition from unicellular organisms to multicellular animals [79]. Until recently, tyrosine kinases were solely recognized in the genomes of Metazoa. Recent genomic analyses have shown phosphotyrosine signaling components present in choanoflagellates, the

closest living unicellular relatives to multicellular animals [80]. The presence of eukaryotic-like tyrosine kinases in unicellular organisms implies that these early signaling molecules may have played a role in the evolution of multicellular organisms in the metazoan lineage [80-83].

Tyrosine Signaling in Choanoflagellates

Choanoflagellates, named after the 'feeding cells' (choanocytes) of sponges, are the nearest ancestors to metazoans (Figure 1-5). *Monosiga brevicollis,* a choanoflagellate species, possesses a spherical cell body 3-10 μm in diameter with a single flagellum, surrounded by a collar of microvilli. Movement occurs through water currents generated by movement of the flagellum (Figure 1-6). More than 125 choanoflagellate species have been identified, and all are unicellular, although some can form simple colonies [80, 84]. Upon comparison of the nuclear and mitochondrial genome between choanoflagellates, sponges, and other metazoans, the presence of metazoan signaling and adhesion gene homologues have been found in the unicellular choanoflagellates [85-88]. The choanoflagellates have a distinct lineage from the metazoans, indicating that the shared genes must come from an earlier ancestor to both groups [89, 90]. Therefore, studying the signaling molecules in choanoflagellate species will give an insight into the evolution of the present-day metazoan counterparts.

Dr. Nicole King and her colleagues have studied the evolution of multicellularity in the choanoflagellate *Salpingoeca rosetta*. *S. rosetta* is a good model species to study the evolution of multicellular organisms due to its ability to exist in both single cell and colonial form. *S. rosetta* can form two different colony types: chain and rosette. The colonies are formed by an incomplete cytokinesis, meaning that the colonies consist of one cloned individual [91]. Neighboring cells are connected by a fine network of intercellular bridges, which allows the movement of small molecules [92]. The ability of the colonial cells to interact with each other in this way makes them especially valuable as a model to study how simple multicellular organisms first evolved. The use of comparative genomics to investigate the ancestral metazoan genome is an important tool for determining which genes are shared by choanoflagellates and metazoans, and to help pinpoint the signaling molecules that were necessary for the transition to multicellularity [80]

Monosiga brevicollis, a member of the choanoflagellates, contains 128 tyrosine kinases in its genome [82]. Besides containing a higher number of TKs as compared to humans (which have 90), the domain architectures of the proteins are also more varied than seen in complex metazoans (Figure 1-7) [82]. Of the 128 tyrosine kinases, 88 are RTKs. Most of these show the typical transmembrane structure of their mammalian counterparts, as well as containing modular domain units. The majority also contain at least one tyrosine in the activation loop of the kinase domain, implying a similar mechanism as seen in Metazoa: increase of activity by phosphorylation in the activation loop. Many domain arrangements seen for the *Monosiga* are not conserved in mammals. As part of this project, we carried out the first biochemical studies of an RTK from a unicellular organism (RTKB2 from *Monosiga* brevicollis). These results will be described in chapter 4 of this thesis.

A wide variety of nRTKs is present in the genome of *M.* brevicollis. *M. brevicollis* contains 40 nRTKs, including orthologues of mammalian nRTKs such as Src, Abl, Tec and Csk, but most of them have no clear counterpart in metazoans. Many of the domain arrangements have never been seen before in Metazoa, such as the linking of an inositol lipid-binding FYVE domain, a phosphotyrosine binding (PTB) domain, or a lipid-binding C2 domain to a kinase domain. The presence of these previously unseen domain combinations reinforces the idea that tyrosine kinases evolved before the split between choanoflagellates and metazoans, and strongly suggests a domain shuffling method of protein evolution [82].

The first time an active tyrosine kinase was observed outside metazoans was in the cell lysates of *Monosiga brevicollis* cells. When lysates were probed with general anti-phosphotyrosine antibodies, 15-18 proteins were detected under starvation conditions. When the *Monosiga* culture was grown under nutrient-rich conditions, changes in the pattern of tyrosine phosphorylation were seen. Additionally, when the cultures were treated with tyrosine kinase inhibitors, cell growth decreased, pointing to an essential role of pTyr signaling in choanoflagellates [86]. Since many of the RTKs were not similar to those in metazoa and the *Monosiga* ligands are unknown, it was not possible to isolate the specific TKs that were activated and deactivated upon environmental change.

Prior to the work discussed in this thesis, the only unicellular nRTKs to be cloned, expressed and studied were from two choanoflagellate species: *Monosiga ovata* and *Monosiga brevicollis*.

Three Src-like kinases and one Csk kinase were identified from the genome of *Monosiga ovata*, a choanoflagellate of a distant clade from *Monosiga brevicollis*. The three Src kinases had varying levels of activity based on Western blotting. Although Csk was able to phosphorylate all isoforms at their negative regulatory site, they still showed significant activity even in the phosphorylated form. Analysis of chimera mutants showed that the lack of stable negative regulation is due to structural changes in the kinase domain of the choanoflagellate Src kinases [87]. These findings suggest that the conserved phosphorylation of the Src C-terminal tail Tyr by Csk was already in existence in *Monosiga ovata*, but the concurrent Src down-regulation that is necessary in multicellular animals for proper functioning was not yet honed.

In *Monosiga brevicollis*, four Src kinase homologues were observed (MbSrc1-4), as well as a Csk homologue (MbCsk) [86]. The isolated SH2, SH3 and kinase domains of MbSrc1 functioned similarly to their counterparts in mammalian Src kinase, with slightly different substrate specificities. Even so, MbSrc1 could functionally replace mammalian Src kinases in functional assays. The regulatory characteristics of MbSrc1 and c-Src were different. In contrast to c-Src, MbSrc1 did not show increased phosphorylation of substrates containing SH2 ligands. The SH2 domain of MbSrc1 did not target the enzyme to phosphorylate synthetic peptides containing SH2 ligands more efficiently than control peptides. Also, the incubation of MbSrc1 with ATP or peptide ligands for SH2 or SH3 did not increase kinase activity. These experiments indicate that the form of MbSrc1 used was not autoinhibited. As seen in *M. ovata*, MbCsk was able to phosphorylate MbSrc1's C terminus, but no inhibition was seen. These results imply that both the functional coupling of the SH2 and kinase domains, as well as Csk-imposed negative regulation of Src did not exist in the choanoflagellates studied, and that these additions may correlate with the transition to multicellular animals [93].

MbSrc4 from *Monosiga brevicollis* was also cloned, expressed, and studied in biochemical assays. This protein is an example of a choanoflagellate tyrosine kinase with a combination of signaling domains not seen in any metazoan. MbSrc4 contains a lipid-binding C2 domain at the N-terminal, followed by the canonical SH3-SH2-kinae domains of metazoan Src kinases. In Western blotting and *in vitro* assays, MbSrc4 was a highly active tyrosine kinase. Its C2, SH3, and SH2 domains had similar targeting functions as in mammalians. The C2 domain exhibited similar membrane-binding activity as the myristoylation sequence in c-Src, most likely due to

convergent evolution. Although the C2 domain could target the protein to the membrane, MbSrc4 could not replace mammalian Src in functional assays. MbSrc4 is not properly targeted to mammalian Src substrates, suggesting a different role in *M. brevicollis* cells [94].

Other Unicellular Metazoan Ancestors

Capsaspora owczarzaki is a filose amoeboid symbiont of the pulmonate snail *Biomphalaria glabrata* (Figure 1-6). Phylogenetic analyses have determined that *Capsaspora* is another unicellular lineage that branches as a sister-group to Metazoa (Figure 1-5) [76, 95-97]. Although closely related to choanoflagellates, *Capsaspora* shares a distinct clade (Filasterea) with the unicellular organism *Ministeria vibrans* (Figures 1-5, 1-6) [98]. *Capsaspora* was initially discovered due to interest in its host snail *B. glabrata*, an intermediate host for the krematode *Schistosoma mansoni*, a human pathogen. The World Health Organization reports that 600 million people are currently at risk for it, mostly in developing countries (WHO Expert Committee, 1993). *Capsaspora* has been found to kill sporocytes from the pathogenic worm, making it a target for medical research [99].

*Capsaspora'*s location on the phylogenetic tree has increased interest in this organism, and studying its genome can help elucidate the evolutionary history of essential genes for multicellularity and the development of Metazoa. *C. owczarzaki* contains 103 putative TK-encoding genes, 92 of them classified as RTKs in 19 different families, and the other 11 nRTKs divided into 8 families. There is great divergence in the RTK architecture, although many of the modular extracellular domains themselves are conserved in Metazoa, where they serve protein interaction functions, such as binding to extracellular ligands or other receptors [98]. The genome of *C. owczarzaki* contains many homologues to nRTKs seen in Metazoa, for example, Src, Csk, Abl, Fak, etc. In chapter 3 I discuss my studies of two Src-like and one Csk-like kinases (CoSrc1, CoSrc2 and CoCsk) in *Capsaspora*, determining their activity as well as investigating the conserved relationship of Src and Csk.

The Filasterea group, containing *Capsaspora owczarzaki* and *Ministeria vibrans*, are the closest relatives to the Metazoa and Choanoflagellate clades. *M. vibrans* is a free-living, bacteria-eating marine protist with radiating tentacles around a central body (Figure 1-6) [97]. PCR-based surveys of the *M. vibrans* genome revealed seven RTKs and eight nRTKs. These

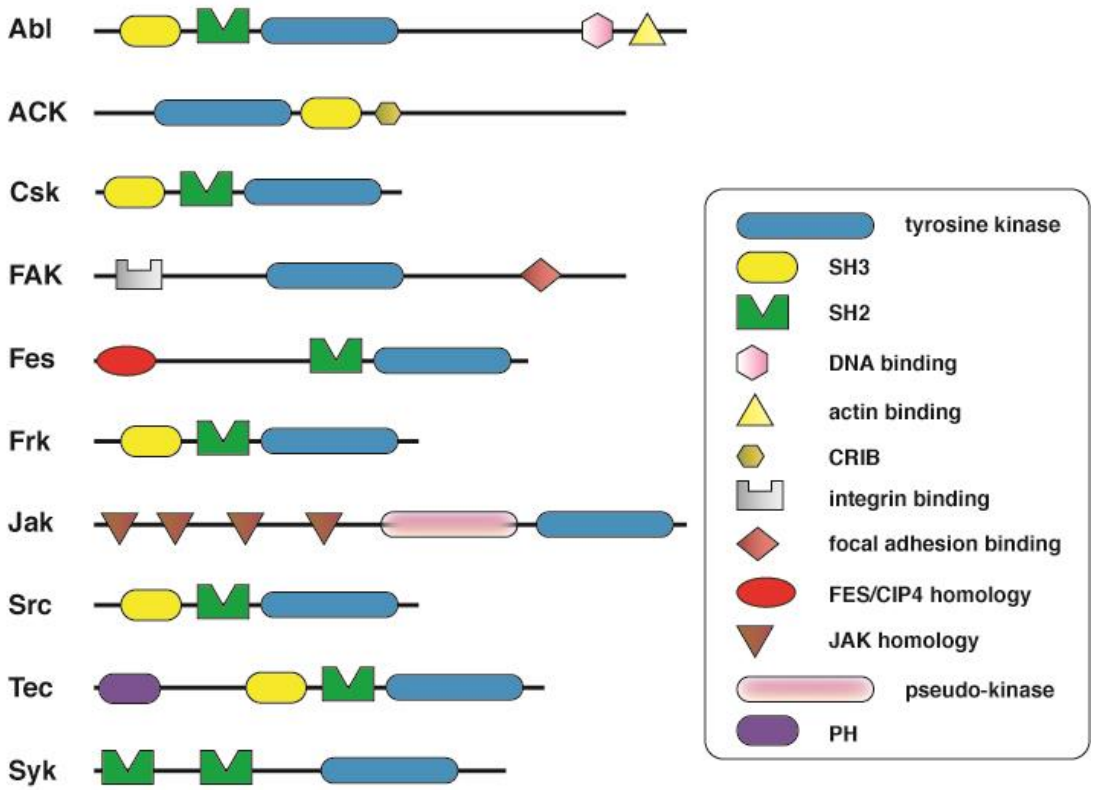
were organized into six nRTK families and one RTK family. The variety of nRTKs is similar to that seen in *Capsaspora*, but the lone RTK family of *M. vibrans* does not share domain architecture with any of the known RTK families, including the RTKs of the closely related *Capsaspora*. Chapter 4 of this thesis shows my experiments with the two Src-like kinases and one Csk kinase in *Ministeria* (MvSrc1, MvSrc2 and MvCsk).

Chapter 5 of this thesis deals with my initial experiments on other signaling components in choanoflagellates. We have focused on identifying substrates for SFKs in choanoflagellates. The first section is focused on studying RTKB2, a receptor tyrosine kinase in the *M. brevicollis* genome. The experiments on RTKB2 are the first biochemical studies to be done on a unicellular receptor tyrosine kinase. The second set of experiments in this chapter examines an *S. rosetta* STAT protein. We have investigated the interaction of this unicellular STAT with *S. rosetta* Src kinase.

Figure 1-1. **Domain architecture of human tyrosine kinases.** (A)Families of non-receptor tyrosine kinases. The most common domain arrangement is SH3-SH2-Kinase domain, as seen in the Src family kinases. (B) Families of receptor tyrosine kinases. The domain architecture of RTKs is not conserved.

Panel B adapted from Institut Europeen de Chimie et Biologie, www.cellbiol.net

A


Abl
ACK
Csk
FAK
Fes
Frk
Jak
Src
Tec
Syk
tyrosine kinase
SH3
SH2
DNA binding
actin binding
CRIB
integrin binding
focal adhesion binding
FES/CIP4 homology
JAK homology
pseudo-kinase
PH

B

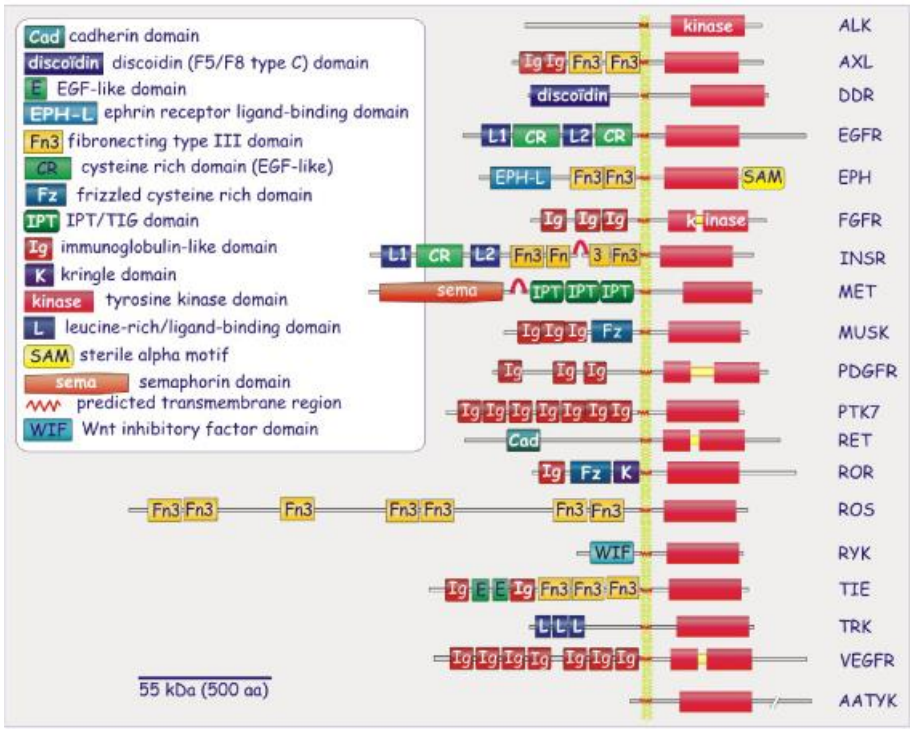
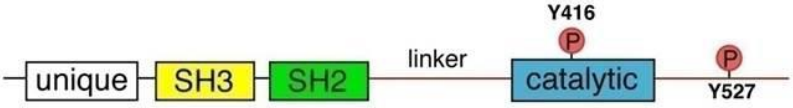

Cad cadherin domain
discoïdin discoidin (F5/F8 type C) domain
E EGF-like domain
EPH-L ephrin receptor ligand-binding domain
Fn3 fibronecting type III domain
CR cysteine rich domain (EGF-like)
Fz frizzled cysteine rich domain
IPT IPT/TIG domain
Ig immunoglobulin-like domain
K kringle domain
kinase tyrosine kinase domain
L leucine-rich/ligand-binding domain
SAM sterile alpha motif
sema semaphorin domain
predicted transmembrane region
WIF Wnt inhibitory factor domain
55 kDa (500 aa)
ALK
AXL
DDR
EGFR
EPH
FGFR
INSR
MET
MUSK
PDGFR
PTK7
RET
ROR
ROS
RYK
TIE
TRK
VEGFR
AATYK

Figure 1-2. **Structure of the Src family kinase Hck.** The SH3 domain is depicted in yellow, the SH2 domain in green, and the catalytic kinase domain is blue. Phosphorylation sites at Tyr-416 (activation loop) and Tyr-527 (C-terminal tail) are highlighted.

Reproduced from Miller, WT (2003) (25).


Y416
P
unique
SH3
SH2
linker
catalytic
P
Y527

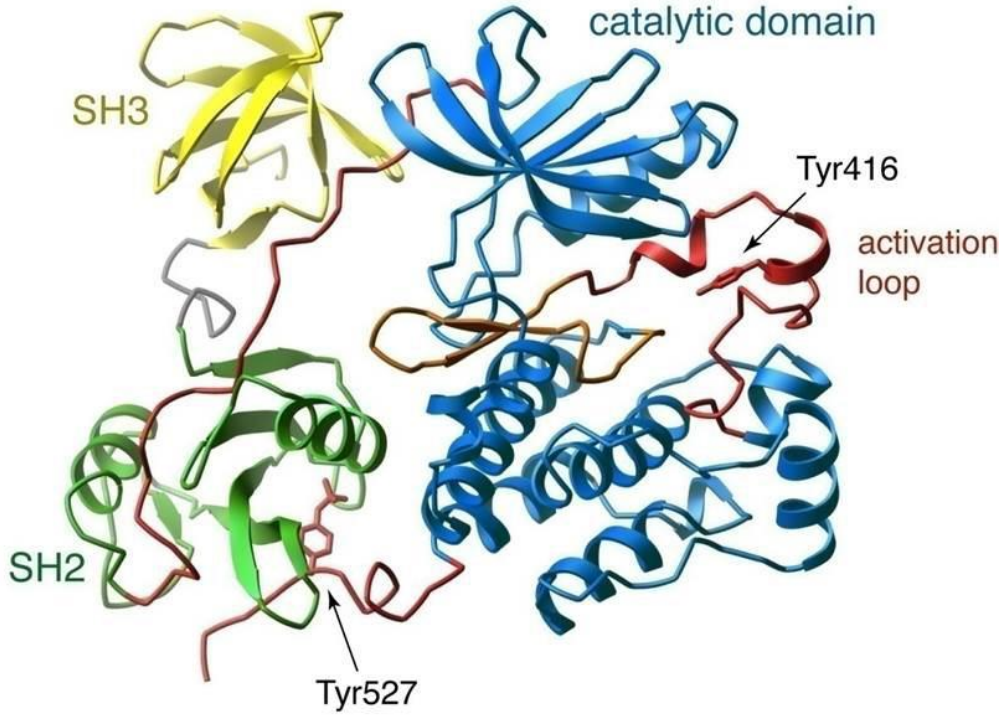

catalytic domain
SH3
Tyr416
activation
loop
SH2
Tyr527

Figure 1-3. **Mechanisms for Src kinase activation.** The inactive form of the Src kinase is shown being activated in three ways: dephosphorylation in the C-terminal tail Tyr-527, SH2 ligation, and SH3 ligation. pY denotes a phosphorylated tyrosine.

From Miller, W.T. (2003) (25)

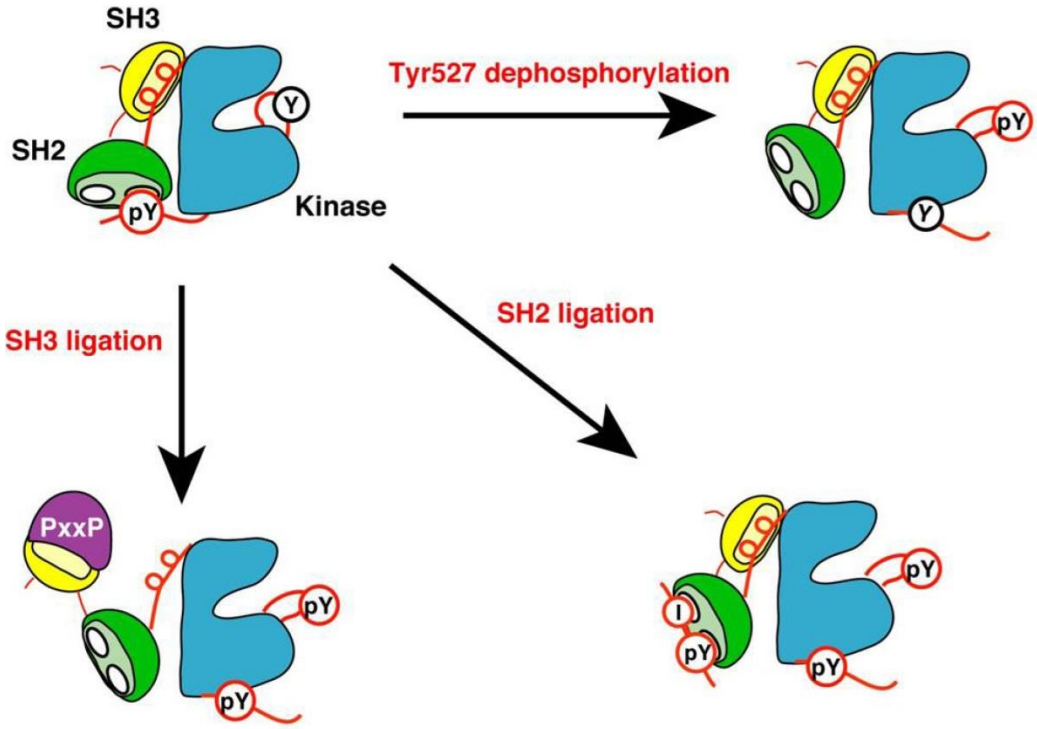
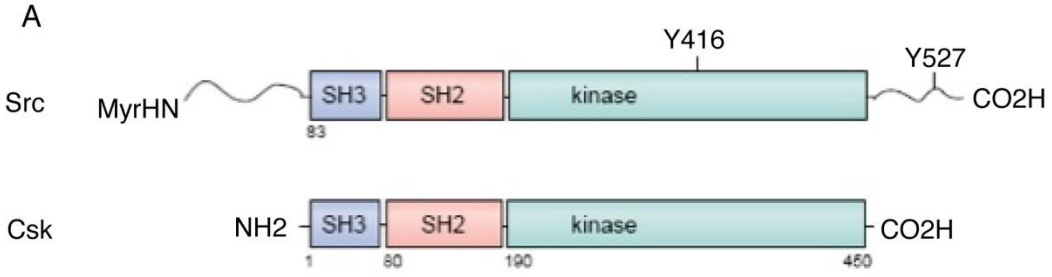

SH3
SH2
Y
pY
Kinase
Tyr527 dephosphorylation
pY
Y
SH3 ligation
SH2 ligation
PxxP
pY
pY
I
pY
pY
pY

Figure 1-4. **Csk domain architecture. (A)** The domain arrangement of Csk is similar to the Src family kinases: an SH3 domain, followed by an SH2 domain and a catalytic kinase domain. Unlike Src, Csk does not contain a myristoylation sequence at the N-terminus, nor the two conserved tyrosines important for regulation: Tyr-416 and Tyr-527. (B) The schematic representation of the crystal structure of Csk. The SH3-SH2 linker and the SH2-kinase linker are both in direct contact with the N-terminal lobe of the catalytic domain. The transition of the α-helix-C from the active to inactive form is shown by a 60° rotation. The substrate docking site and the position of Ser-364 are shown in the kinase domain.

Panel (B) reproduced from Chong et al. (2005) (53)

A


Y416
Y527
Src
MyrHN
SH3
SH2
kinase
CO2H
83
Csk
NH2
SH3
SH2
kinase
CO2H
1
80
190
450

B

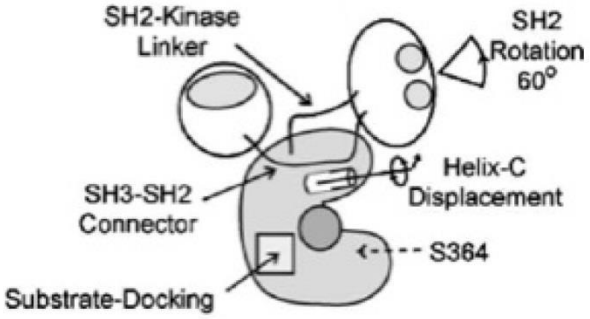

SH2-Kinase
Linker
SH2
Rotation
60°
Helix-C
Displacement
SH3-SH2
Connector
S364
Substrate-Docking

Figure 1-5. **Phylogenetic tree depicting the metazoan ancestral lineage.** The phylogenetic tree shows the positions of the choanoflagellates *M. brevicollis* and *S. rosetta*. *Ministeria vibrans* and *Capsaspora owczarzaki* form a closely-related clade, the Filasterea.

Figure from Dr. Nicole King, University of California at Berkeley

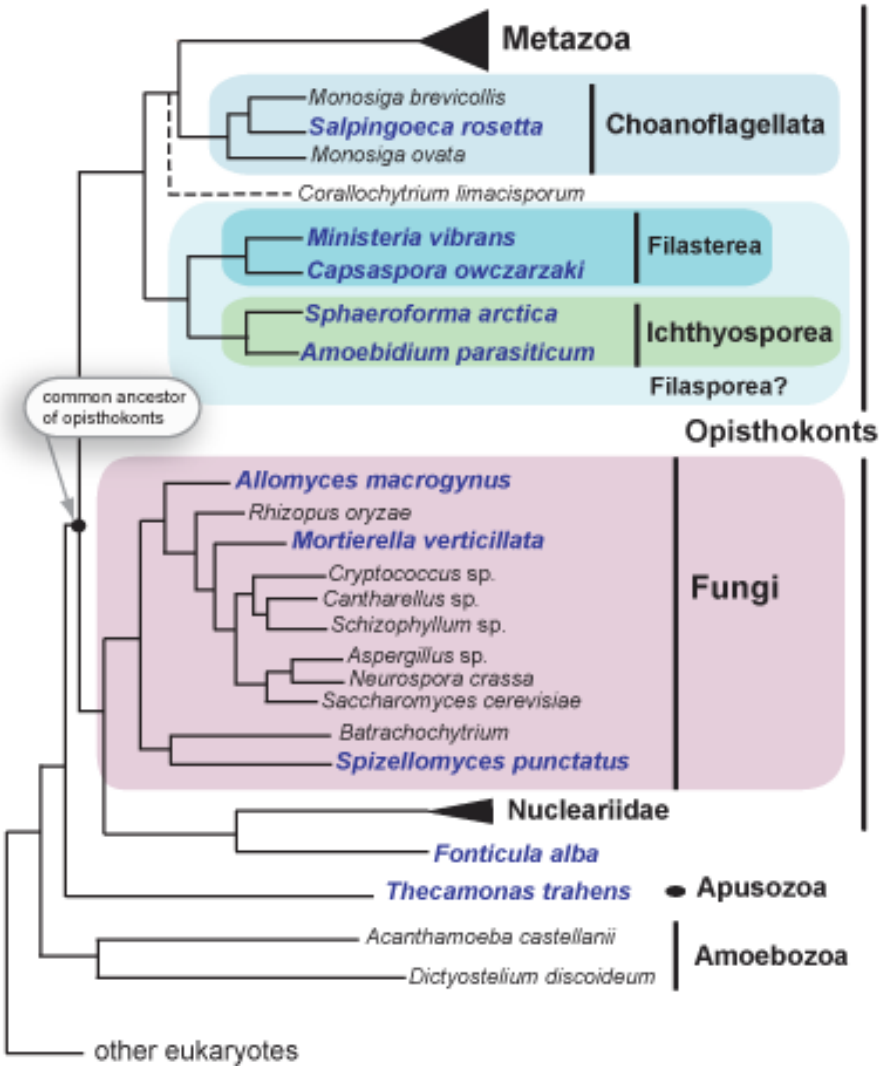
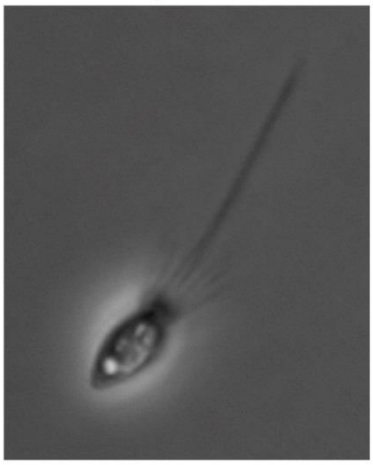

Metazoa
Monosiga brevicollis
Salpingoeca rosetta
Monosiga ovata
Choanoflagellata
Corallochytrium limacisporum
Ministeria vibrans
Capsaspora owczarzaki
Filasterea
Sphaeroforma arctica
Amoebidium parasiticum
Ichthyosporea
Filasporea?
common ancestor of opisthokonts
Opisthokonts
Allomyces macrogynus
Rhizopus oryzae
Mortierella verticillata
Cryptococcus sp.
Cantharellus sp.
Schizophyllum sp.
Aspergillus sp.
Neurospora crassa
Saccharomyces cerevisiae
Batrachochytrium
Spizellomyces punctatus
Fungi
Nucleariidae
Fonticula alba
Thecamonas trahens
Apusozoa
Acanthamoeba castellanii
Dictyostelium discoideum
Amoebozoa
other eukaryotes

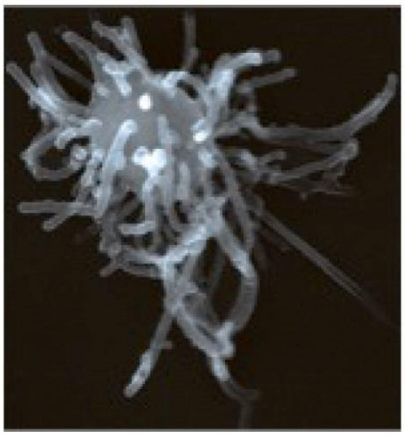
Figure 1-6. **Images of *Monosiga brevicollis, Capsaspora owczarzaki*, and *Ministeria vibrans*.**

Images from Mark Dayel (University of California at Berkeley), Arnau Sebe-Pedros (University of Barcelona), and Inaki Ruiz-Trillo (University of Barcelona), respectively.

Monosiga Brevicollis



Capsaspora owczarzaki



Ministeria vibrans

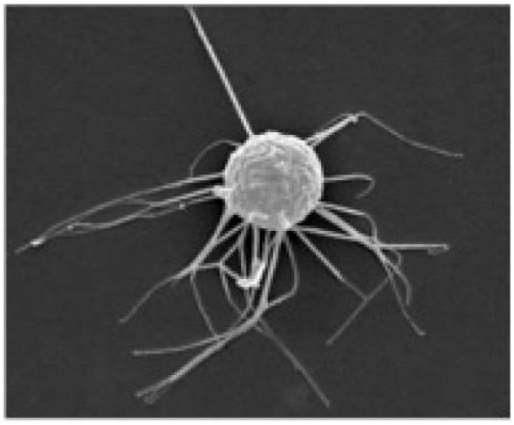
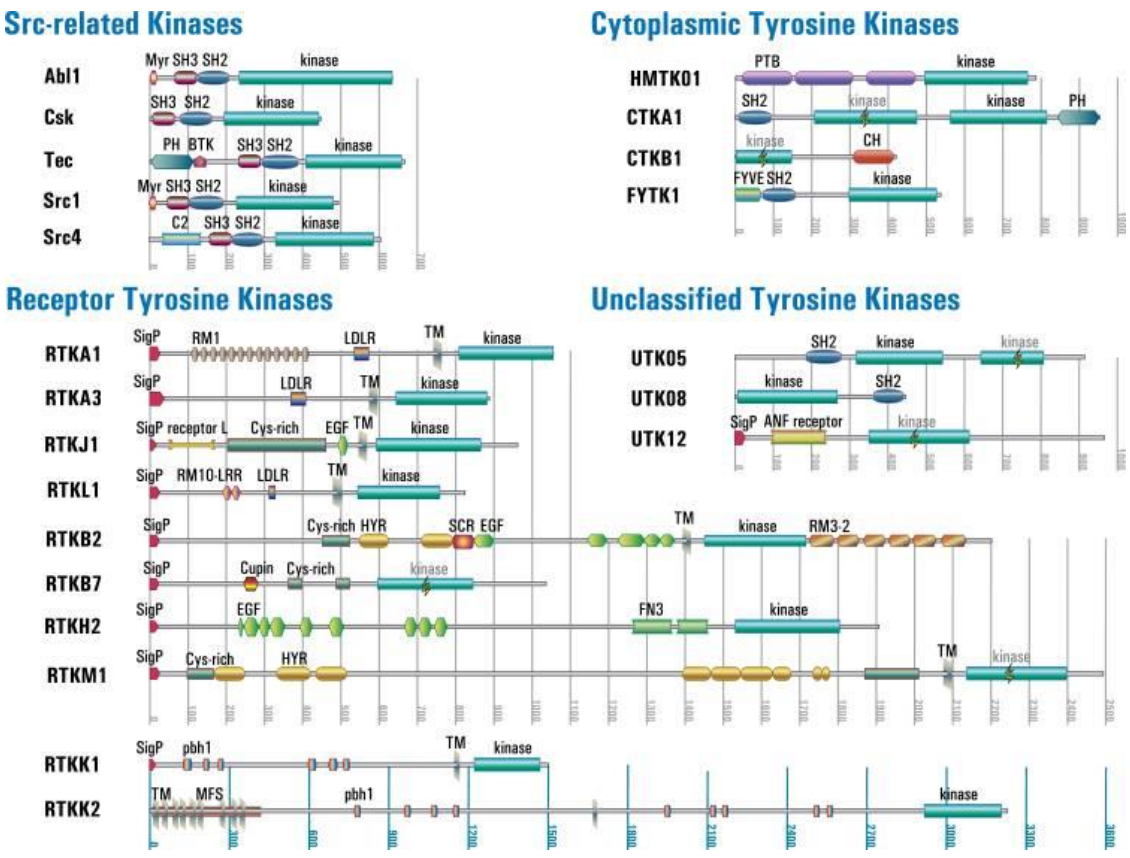


Figure 1-7. **Domain architecture of representative tyrosine kinases from *Monosiga brevicollis.*** The *M. brevicollis* genome contains 128 tyrosine kinases. The RTKs show mainly typical domain units. The nRTKs are more varied. Although there are mammalian homologs of Src, Abl, Tec, and Csk, most lack any clear counterpart in metazoans. Predicted inactive kinase domains are indicated by a lightning bolt, partial matches to domains are represented by shortened icons.
Figure from Manning et al, (2008) (73)


Src-related Kinases
Abl1
Csk
Tec
Src1
Src4
Myr SH3 SH2
kinase
SH3 SH2
PH BTK
C2
Cytoplasmic Tyrosine Kinases
HMTK01
CTKA1
CTKB1
FYTK1
PTB
SH2
CH
FYVE
PH
Receptor Tyrosine Kinases
RTKA1
RTKA3
RTKJ1
RTKL1
RTKB2
RTKB7
RTKH2
RTKM1
RTKK1
RTKK2
SigP
RM1
LDLR
TM
receptor L
Cys-rich
EGF
RM10-LRR
HYR
SCR
Cupin
FN3
RM3-2
pbh1
MFS
Unclassified Tyrosine Kinases
UTK05
UTK08
UTK12
ANF receptor

Chapter 2

Materials and Methods

Materials and Methods

Reagents and antibodies

Nickel-nitriloacetic acid resin was purchased from Qiagen. Anti-phosphotyrosine antibody (clone 4G10) was from Millipore. Anti-FLAG antibody, EZview Red Anti-FLAG M2 Affinity Gel, CNBR-linked Sepharose, leupeptin, aprotinin, PMSF, sodium vanadate, bovine serum albumin (BSA) and ethanolamine were all bought from Sigma. Anti-V5 antibody and glutathione-agarose-linked beads were from Invitrogen. DMEM, Trypsin-EDTA, SF900II-SFM, Penicillin, Streptomycin were from GIBCO (Cellgro). HRP conjugated secondary anti-mouse and anti-rabbit antibodies, as well as ECL and ECL+ kits, were from Amersham.

Cell Culture

Mammalian Src/Yes/Fyn deficient cells (SYF) were maintained in DMEM (Cellgro, Mediatech, Inc) with 10% fetal bovine serum (Sigma) 1X antibiotic/antimycotic at 37°C in 5% CO_2. Sf9 insect cells were maintained in Sf-900 medium (Gibco) supplemented with 5% fetal bovine serum and 1X antibiotic/antimycotic (Cellgro, Mediatech, Inc). A culture of *C. owczarzaki* ATCC30864 was purchased from the American Type Culture Collection (ATCC) and maintained at 23°C in ATCC medium 1034.

Cloning and site-directed mutagenesis

C. owczarzaki CoSrc1, CoSrc2 and CoCsk, as well as *M. vibrans* MvSrc1, MvSrc2, and MvCsk were predicted in the draft genome sequences, which are available at http://www.broadinstitute.org/annotation/genome/multicellularity_project/, by the use of Augustus. Model parameters of the prediction were specifically optimized for each organism. The cDNA's were amplified and cloned from a *Capsaspora* or *Ministeria* cDNA library by polymerase chain reaction with specific primers by Hiroshi Suga and Inaki Ruiz-Trillo

(University of Barcelona). We confirmed that the sequences have no errors that change amino acids. For expression in insect cells, CoSrc1 and CoSrc2 (residues 111-566 and 135-590, respectively) were cloned into the BamHI and XbaI sites of pFastbac-HtC (Invitrogen). MvSrc1 and MvSrc2 were also cloned into the BamHI and XbaI sites of pFastbac, inserting residues 95-555 and 79-596, respectively.

For mammalian expression, N-terminally FLAG-tagged full-length CoSrc1 and CoSrc2 (566 and 590 residues long, respectively), and N-terminally FLAG-tagged full-length MvSrc1 and MvSrc2 (555 and 596 residues) were subcloned into the XbaI and BamHI restriction sites of p3x-FLAG-CMV (Sigma).

The cDNA for CoCsk (encoding a protein of 501 residues) and MvCsk (456 residues) were subcloned into pGex-4T-1 (GE Healthcare) using EcoRI and XhoI restriction sites. V5-tagged CoCsk and MvCsk were expressed in mammalian cells after subcloning into the BamHI and XbaI sites of pEF1/V5-HisA (Invitrogen).

CoSrc1 and CoSrc2 activation loop mutants were produced by site-directed mutagenesis using the Stratagene Quikchange kit, following manufacturer's instructions. His-tagged CoSrc1 and CoSrc2 constructs were mutated at their activation loop tyrosine to phenylalanine (Y447F and Y471F, respectively). GST-tagged CoCsk and MvCsk constructs were also altered by site-directed mutagenesis. A glycine in the kinase-SH2 interface of CoCsk and MvCsk was mutated (G278N and G226N, respectively).

Amino acid sequences were aligned using ClustalW and formatted by BOX-SHADE (version 3.3.1 by K. Hofmann and M. D. Baron) in UCSK workbench 3.2. Models of 3-dimensional protein structures were done in Pymol.

The protein sequence of RTKB2 was predicted from release 1.0 of the *Monosiga brevicollis* genome. The kinase domain (residues 1450-1724) or C-terminal tail (residues 1722-2200) were amplified by PCR from an *M. brevicollis* cDNA library. For baculovirus expression, the kinase cDNA was cloned into the BamHI and XbaI sites of pFastbac-Htb (Invitrogen). For mammalian cell expression, the C-terminal tail cDNA was cloned into the EcoRI and BamHI sites of plasmid pEGFP-C1 (Clontech). For bacterial expression of RM2-6, the gene was synthesized with a sequence optimized for *E. coli* (GenScript) and cloned into plasmid pET-15b.

. The *S. rosetta* Src and STAT peptides were a gift from Tera Levin and Nicole King (University of California, Berkeley), as were *S. rosetta* Src in p3xFLAG-CMV (Sigma) and *S. rosetta* STAT in pEF1/V5-HisA (Invitrogen).

Cell transfection and Western blotting

Cells were plated at a density of $1x10^6$ cells per 100 mm plate. At 50% confluency (24 hours later), 1-2 µg of DNA constructs were transfected with TransIT LT1transfection reagent (Mirus) at a ratio of 1:3 (DNA:TransIT). Cells were harvested after 48 hours and lysed in RIPA buffer (50 mM Tris-HCl [pH 7.4], 150 mM NaCl, 1 mM EDTA [pH8], 1 mM phenylmethanesulfonyl fluoride, 1% NP-40, 0.25% sodium deoxycholate, 1 mM Na_3VO_4, 1 mg/mL aprotinin, and 1 mg/mL leupeptin) for 1 hour at 4°C. Lysates were centrifuged at 14000*g* for 15 minutes at 4°C, separated by electrophoresis in 10% SDS-PAGE, transferred to PVDF membrane, and probed with antiphosphotyrosine, anti-FLAG, and anti-V5 antibodies for CoSrc and MvSrc, or antiphosphotyrosine, anti-GFP, and anti-FLAG for RTKB2-CT.

Tyrosine kinase assays

CoSrc1, CoSrc2, MvSrc1, and MvSrc2 activity assays were conducted with [γ-^{32}P]ATP using a phosphocellulose paper binding assay [100]. Reactions were performed in 20 mM Tris-HCl (pH 7.4), 10 mM $MgCl_2$, 0.25 mM ATP, varying concentrations of peptide substrate, and [γ-^{32}P]ATP (100-500 cpm/pmol). Peptide substrates were purified by reverse-phase high-pressure liquid chromatography. The substrate specificity reactions were performed with the following peptides: Src peptide, AEEEIYGEFEAKKKKG; EGFR peptide, AEEEEYFELVAKKKG; Abl peptide, EAIYAAPFAKKKG; and Kemptide, LRRASLG. Substrate targeting experiments were performed using the following peptides: SH2 substrate, RRLEDAIYAAGGGGGEPPQpYEEIG; SH2 control, RRLEDAIYAAGGGGGEPPQFEEIG; SH3 substrate, AEEEIYGEFGGRGAAPPPPPVPRGRG; and SH3 control, AEEEIYGEFGGRGAAAAAAAVPRGRG. To study the autophosphorylation-dependent activity of CoSrc1 and CoSrc2, the enzymes were first treated with immobilized GST-tagged

Yersinia tyrosine phosphatase (YOP) for 30 minutes at 30°C, followed by incubation with 0.25 mM ATP in kinase assay buffer for 30 minutes at 30°C. For mass spectrometry analysis, CoSrc (50 pmol) was treated with YOP, allowed to autophosphorylate, then added to 0.1 M ammonium bicarbonate, 0.1 M DTT, and 0.2 M iodoacetamide. The sample was then digested overnight with trypsin. Results were analyzed on a Thermo Fisher Scientific LTQ Orbitrap CL ETD instrument. To detect the specific phosphorylation of serine in the Kemptide synthetic peptide, MALDI-TOF was also conducted on the peptide after phosphorylation by MvSrc2 as described above.

CoCsk and MvCsk were assayed with [γ-^{32}P]ATP and poly(Glu_4-Tyr). The reactions were stopped at 2, 8, 15, and 30 minutes by spotting onto Whatman 3MM paper and washing with 5% trichloroacetic acid at 55°C, followed by liquid scintillation counting. CoCsk and MvCsk activity towards CoSrc1/2 and MvSrc1/2, respectively, was tested by incubating the proteins in kinase buffer with [γ-^{32}P]ATP. The reactions were stopped by addition of Laemmli buffer and the mixtures analyzed by 10% SDS-PAGE and autoradiography. To quantify phosphorylation, bands were excised, dissolved in hydrogen peroxide, and analyzed by scintillation counting.

RTKB2 activities against synthetic peptides were analyzed using the phosphocellulose paper binding assay as described above. The sequences of the peptides tested were: Src peptide, AEEEIYGEFEAKKKKG; RTKB2-1, SEEVYGAVVDKKK; RTKB2-2, AEEVYEAIADKKK; insulin receptor substrate (E4YM4), KKEEEEYMMMMG. Kinase assays for *S. rosetta* were performed using the Src substrate peptide (as above) or *S. rosetta* STAT peptide, GEAKNSKLLTKGYIPA.

Protein purifications

The pFastbac-HtC vector encoding CoSrc1, CoSrc2, MvSrc1, MvSrc2 or RTKB2 was used to transfect Sf9 cells using the Bac-to-Bac system (Invitrogen). Baculovirus stocks were produced by transfecting Sf9 cells. After two rounds of viral amplification, approximately 30 mL of virus was used to infect 400 mL of Sf9 cells at $2x10^6$ cell/mL. After 72 hours, infected

cells were collected by centrifugation, and resuspended in lysis buffer [50 mM Tris-HCl (pH 8.0, 5 mM 2-mercaptoethanol, 100 mM KCl, 1 mM PMSF, 1% Nonidet P-40, 1 mg/mL aprotinin, 1 mg/mL leupeptin]. They lysates were French pressed twice, and cell debris removed by centrifugation at 40,000*g* for 30 minutes, filtered with a 0.8 μm filter and added to a 5 mL Ni-NTA column (Qiagen). The column was washed with Buffer A [50 mL of 20 mM Tris-HCl (pH 8.0, 500 mM KCl, 20 mM imidazole, 5 mM 2-mercaptoethanol, 10% glycerol], and further washed with 25 mL of Buffer B [20 mM Tris-HCl (pH 8.0, 1 M KCl, 5 mM 2-mercaptoethanol, 10% glycerol]. The column is washed again with 25 mL of Buffer A, and then eluted with 20 mL of Buffer C [20 mM Tris-HCl (pH 8.0), 100 mM KCl, 100 mM imidazole, 5 mM 2-mercaptoethanol, 10% glycerol). All washes and elutions were carried out at 4°C. The 1 mL containing the His_6-tagged proteins were pooled, concentrated, and stored in 40% glycerol at -20°C. His-tagged RM2-6 was expressed in 1-liter cultures of *E. coli* and purified by Ni-NTA chromatography as well.

CoCsk and MvCsk were expressed as GST-tagged proteins in *E. coli* BL21 cells. Expression was induced with 0.25 mM IPTG for 5 hours at 30°. Cells were collected by centrifugation and lysed through French pressing in Lysis Buffer [50 mM Hepes buffer (pH 7.4), 100 mM NaCl, 100 mM EDTA, 1% Triton X-100, 10% glycerol, 5 mg/mL aprotinin, 5 mg/mL leupeptin, and 1 mM PMSF]. Lysates were centrifuged at 11,000 *g* for 30 minutes. The supernatant was rocked for 2 hours with 10 mL of glutathione-agarose beads equilibriated in 50 mM Hepes buffer (pH 7.4) containing 100 mM EDTA. The beads were washed twice with 10 mL of Lysis Buffer, and then further washed 4 times in 50 mL of 50 mM Hepes buffer (pH 7.4) with 100 mM EDTA and 100 mM NaCl, followed by a last wash in 50 mL of 50 mM Tris (pH 8.0). The beads were then loaded onto a column and eluted with 20 mL of 50 mM Tris (pH 8.0) with 20 mM glutathione in 20 1 mL fractions. Fractions containing GST-tagged proteins were pooled, concentrated, and stored in 40% glycerol at -20°C.

Binding assays

Interactions between CoSrc1/CoSrc2 with CoCsk were examined through binding assays. FLAG-tagged CoSrc1/2 were overexpressed in SYF cells, harvested and incubated in lysis buffer for 1 hour. GST-CoCsk (or GST control) was immobilized on glutathione-agarose beads. Lysates (0.5 mg) expressing CoSrc1 or CoSrc2 were incubated with the resins for 1 hour at 4°C. After binding, the beads were washed 3 times with 50 mM Hepes buffer (pH 7.4), 100 mM NaCl, 100 mM EDTA, 1% Triton X-100, 10% glycerol, 5 mg/mL aprotinin, 5 mg/mL leupeptin, and 1 mM PMSF. Bound proteins were eluted with Laemmli buffer and separated by 10% SDS-PAGE. The proteins were transferred to a PVDF membrane and analyzed by Western blotting with anti-FLAG antibody.

Production and Purification of CoSrc-Specific Antibodies

Specific antibodies recognizing CoSrc1 and CoSrc2 were developed. Peptides derived from their N-terminal unique regions were used to create rabbit antibodies (Genemed Synthesis, Inc.). The peptides used to generate antibodies toward CoSrc1 were CQAQQPMLPGQIMAQQ and CSTLPGQRPGGPGGRV; the peptides used for CoSrc2 were CSNSKPHDPHDSDFKVSPSG and CSTTFTAPTPSTNSLK. These antibodies were then isolated from sera through affinity purification. To prepare affinity matrices, synthetic peptides (corresponding to the sequences used to generate antibodies) were added at a concentration of 0.2 mg/mL to 0.8 grams of CNBr-Sepharose that had been swelled for 10 minutes in 1 mM HCl. These reactions proceeded for 1 hour at room temperature and then at 4°C overnight. Beads were washed, incubated with 100 mM ethanolamine for 1 hour at room temperature, washed again, and stored in 500 mM NaCl in PBS; 1-2 mL of affinity matrices was incubated with 20 mL of serum containing CoSrc1 or CoSrc2 antibodies and 30 mL of phosphate-buffered saline overnight at 4°C. Beads were washed with 500 mM NaCl in phosphate-buffered saline, transferred to a column, and eluted with 0.2 M acetic acid (pH 2.7) and 500 mM NaCl in a volume of 10-20 mL. Antibodies were stored in 50% glycerol.

Localization experiments

Localization experiments were conducted by Hiroshi Suga and Inaki Ruiz-Trillo (University of Barcelona). Cells were grown on a coverslip and fixed with 4% paraformaldehyde freshly prepared in an ATCC medium 1034 phosphate buffer solution (pH 6.5) for 10 minutes. Cells were incubated with rabbit anti-CoSrc antibodies (1:10 concentration), together with mouse anti-α-tubulin (Sigma; 1:500), overnight at 4°C. Secondary antibodies [anti-mouse Alexa Fluor 488 and anti-rabbit Alexa Flour 568 (Invitrogen) in a 1:500 ratio) were then applied after washing, and cells were incubated for 1 hour at room temperature. After the secondary antibodies had been washed off, Hoechst 33342 (Sigma) was applied and washed two more times. Cells were then mounted with fluorescence mounting medium (Dako) and observed under a TCS SP5 confocal microscope (Leica). All these procedures were conducted on a coverslip, exchanging liquid gently by pipetting. Cells were kept attached on the glass during the staining and microscopy. Images were processed with ImageJ.

Immunoprecipitation assays

FLAG-tagged CoSrc1, CoSrc2, MvSrc1 and MvSrc2 were expressed in mammalian SYF cells alone, or together with CoCsk or MvCsk. Cells were lysed as described above, and CoSrc1/2 and MvSrc1/2 were isolated from lysates (1 mg) by incubation with 40 µL of anti-FLAG affinity gel for 2 hours at 4°C. Samples were centrifuged at 5000*g* for 10 minutes at 4°C. After being washed with phosphate-buffered saline, the protein-bound beads were used to measure Src activity toward a synthetic Src substrate peptide with the phosphocellulose binding assay.

The CoSrc1- and CoSrc2-specific antibodies described above were used to immunoprecipitate CoSrc1 and CoSrc2 from SYF or *C. owczarzaki* cell lysates. CoSrc1/2 were overexpressed in YSF cells, and endogenous levels of the protein were used from *Capsaspora* cells. Cells were lysed in RIPA buffer for 1 hour at 4°C and then centrifuged for 10 minutes at the same temperature. Lysates (1 mg) were incubated with CoSrc1- and CoSrc2-specific antibodies for 1 hour at 4°C, followed by incubation with immobilized Protein A beads for an

additional 2 hours at 4°C. The samples were centrifuged, washed with PBS, incubated with peptide substrate and [γ-^{32}P]ATP, and assayed using the phosphocellulose binding assay.

For immunoprecipitation assays with RTKB2, lysates (1 mg) were pre-cleared by mixing with 50 mL of protein A-agarose in lysis buffer for one hour at 4°C. After pre-clearing, 2 µg of anti-GFP antibody was added to the lysate and incubated for one hour at 4°C with rocking. Antibody protein complexes were collected with 50 µL gel protein A beads. The beads were washed 5 times in lysis buffer and boiled in 40 µL gel loading buffer. After separation by SDS-PAGE, proteins were transferred to PVDF membrane and analyzed by Western blotting. Western blotting experiments were carried out with the following antibodies: antiphosphotyrosine antibody, anti-FLAG antibody, and anti-GFP antibody. Detection was enhanced by chemiluminescence.

V5-tagged *S. rosetta* STAT was co-expressed with c-Src or *S. rosetta* Src in SYF cells. Cells were lysed in RIPA buffer, and lysates (1 mg) were incubated with anti-V5 antibody for immunoprecipitation. Antibody protein complexes were collected with 50 µL of protein A beads and washed as above. Proteins were separated by SDS-PAGE, transferred to PVDF membrane and analyzed by Western blotting. Membrane was probed with antiphosphotyrosine antibody, anti-V5 antibody, and anti-FLAG antibody.

Immunofluorescence microscopy

Cells expressing GFP-RTKB2 were grown on 35 mm glass bottom dishes (In Vitro Scientific). The cells were first washed with 1x PBS, then fixed with dilute 3.7% formaldehyde in 1x PBS for 15 minutes at room temperature. The fixed cells were washed several times with 1x PBS then mounted on coverslips using VECTASHIELD medium (Vector Laboratories). The GFP expressing cells were visualized by epifluorescence microscopy using a Aeiss Axiovert 200M inverted microscope and a Plan Apochromat 63x/1.40 oil objective. Images were captured using a GFP-Chroma Filter Set, AxioVision software and an AxioCam MRm CCD camera.

.

Chapter 3

Lack of Csk-mediated Src regulation in *Capsaspora owczarzaki**

*The contents of this chapter have been published in Schultheiss et al., Biochemistry, **51**, 8267-77 (2012)

Abstract

Discovery that unicellular choanoflagellates (the closest phylogenetic group to metazoans) possess numbers of tyrosine kinases comparable to those in complex metazoans shattered the long-held idea that these signaling systems arose spontaneously in multicellular animals. To further study the development of tyrosine kinases, we went back one step further in the evolutionary tree: the unicellular *Capsaspora owczarzaki*. *C. owczarzaki* is a unicellular protist representing the sister group to choanoflagellates and metazoans. Two Src-like kinases have been identified in its genome (CoSrc1 and CoSrc2), both of which have the arrangement of SH3, SH2, and catalytic domains seen in mammalian Src kinases. In *Capsaspora* cells, CoSrc1 and CoSrc2 localize to punctate structures in filopodia that may represent primordial focal adhesions. We have cloned, expressed, and purified both enzymes. CoSrc1 and CSrc2 are active tyrosine kinases. Mammalian Src kinases are normally regulated in a reciprocal fashion by autophosphorylation in the activation loop (which increases activity) and by Csk-mediated phosphorylation of the C-terminal tail (which inhibits activity). Similar to mammalian Src kinases, the enzymatic activities of CoSrc1 and CoSrc2 are increased by autophosphorylation in the activation loop. We have identified a Csk-like kinase (CoCsk) in the genome of *C. owczarzaki*. We cloned, expressed, and purified CoCsk and found that it has no measurable tyrosine kinase activity. Furthermore, CoCsk does not phosphorylate or regulate CoSrc1 or CoSrc2 in cells or *in vitro*, and CoSrc1 and CoSrc2 are active in *Capsaspora* cell lysates. Thus, the function of Csk as a negative regulator of Src family kinases appears to have arisen with the emergence of metazoans.

Introduction

Multicellularity evolved several times in eukaryotes [77, 101]. Multicellular animals first emerged during the Cambrian explosion, 600 million years ago. The development of multicellular metazoans is particularly interesting due to their possession of highly specialized cells that require complex systems for cell-cell interaction and communication. Many protein domains involved in animal cells signaling, including canonical tyrosine kinases such as Src and Csk, are not present in bacteria, plants, or even fungi, the closest multicellular relatives to animals [70]. Many multicellular-specific functions such as cell communication, proliferation, and differentiation are controlled by protein tyrosine kinases, which can be classified into receptor and nonreceptor tyrosine kinases [102, 103]. Receptor tyrosine kinases mainly receive signals through their extracellular domains, and then initiate signal transduction pathways through their intracellular domains. Nonreceptor tyrosine kinases are found within the cytoplasm, acting within cells and transducing the signal initiated by receptors. A wide host of phosphotyrosine-based signaling components exist in unicellular choanoflagellates, the closest living relatives to metazoans [80]. This suggests that the evolution of these ancestral communication and adhesion proteins may have played a role in the transition to multicellularity. Choanoflagellates and Metazoa contain many of the same signaling molecules and domains, and their common ancestor might contain these genes. The closest known relative of these two groups is the filasterean *Capsaspora owczarzaki*, a unicellular protist identified in 1980.

Capsaspora owczarzaki, a unicellular eukaryotic symbiont, was discovered by scientists investigating the resistance of some strains of *Biomphalaria glabrata* snails to infection with *Schistosoma mansoni* (a widespread human pathogen) [104]. Upon genomic analysis, it was found that *C. owczarzaki*, similarly to the choanoflagellates, also has a large and diverse complement of tyrosine kinases. There are 103 tyrosine kinases encoded in the *Capsaspora* genome, 92 of these thought to be receptor tyrosine kinases with an intracellular kinase domain, a transmembrane domain, a signal peptide, and motifs in the extracellular portion. The remaining 11 tyrosine kinases are classified as nonreceptor tyrosine kinases, with orthologues in 6 of the 10 major metazoan cytoplasmic tyrosine kinase: Src, Tec, Csk, Abl, Fak, and Fes [98].

In addition to signaling molecules such as tyrosine kinases, the evolution of animals from unicellular ancestors also required mechanisms for cell adhesion and intercellular communication. A major mechanism for cell adhesion in metazoans is integrin-mediated adhesion and signaling. The integrin adhesion complex mediates interactions between cells and the extracellular matrix. Until recently, this machinery was considered to be metazoan specific. Genomic data showed that *Capsaspora owczarzaki* contains canonical metazoan-type integrin adhesion and signaling machinery with a full set of integrin adhesion complex components [105].

Here we report that *C. owczarzaki* has two Src-like kinases (CoSrc1 and CoSrc2), as well as a Csk homologue (CoCsk). In *Capsaspora* cells, CoSrc1 and CoSrc2 localize to punctate structures in filipodia. Since we know that *Capsaspora* contains the complete machinery for integrin-mediated adhesion, these spots may correspond to primordial focal adhesions. CoSrc1 and CoSrc2 have a domain organization similar to that of their mammalian counterparts. They are both active kinases, and this activity is reliant on autophosphorylation on the activation loop tyrosine. CoSrc1 and CoSrc2 display a similar substrate specificity to mammalian Src, though there is a lack of SH3 substrate targeting, a trait shared with MbSrc1 from the choanoflagellate *M. brevicollis.* Surprisingly, CoCsk displays no activity toward both a general tyrosine kinase substrate or CoSrc, its putative cellular substrate. In *Capsaspora* cells CoSrc1 and CoSrc2 display high levels of activity, supporting the notion that they are not down-regulated by Csk. The lack of Csk-mediated regulation of Src in *Capsaspora* suggests that this highly conserved system arose later at the onset of metazoan evolution. We hypothesize the tight regulation of Src was a necessary development in the evolution of proper cell-cell interaction in multicellular animals.

Results

Cloning CoSrc1 and CoSrc2 from *C. owczarzaki*. Two Src-like kinases were identified in the genome of *C. owczarzaki* , we have designated CoSrc1 and CoSrc2. The *Capsaspora* Src-like kinases possess predicted SH3, SH2, and kinase catalytic domains in an arrangement similar to that of mammalian Src kinases and *Monosiga brevicollis* Src, MbSrc1 (Figure 3-1). A phylogenetic tree is shown in Figure 3-2, depicting the relationship among tyrosine kinases from *C. owczarzaki*, the choanoflagellate *M. brevicollis*, the sponge *Amphimedon queenslandica*, fly, and human. Sequence conservation in CoSrc1 and CoSrc2 is weakest in the N-terminal regions. The two critical regulatory tyrosines are conserved in *Capsaspora* (Figure 3-1): CoSrc1 and CoSrc2 both contain a tyrosine (Tyr-447 and Tyr-441, respectively) corresponding to the site of autophosphorylation in the c-Src activation loop (Y416). Phosphorylation of this residue in mammalian Src kinases increases activity. The conserved tyrosine in the mammalian C-terminal tail (Y527) is also present in CoSrc1 and CoSrc2 (Tyr-558 and Tyr-582, respectively). In mammalian Src kinases, this tyrosine in specifically phosphorylated by Csk, promoting an autoinhibited form of the kinase. In the choanoflagellates *Monosiga ovata* and *M. brevicollis*, Csk homologues phosphorylate the Src C-terminal but fail to repress Src activity ([87, 93].

CoSrc1 and CoSrc2 are Active Tyrosine Kinases. We expressed and purified CoSrc1 and CoSrc2 using the Sf9/baculovirus system. CoSrc1 and CoSrc2 both showed concentration-dependent activity toward a synthetic Src substrate peptide. CoSrc2 was more active at all protein concentrations than CoSrc1 (Figure 3-3A). The specific activities of CoSrc1 and CoSrc2 (0.9×10^{-3} and 1.7×10^{-3} mol min^{-1} mg^{-1}, respectively) were comparable to that of mammalian c-Src assayed in a similar manner (3.0×10^{-3} mol min^{-1} mg^{-1}) [29]. Next, we assayed CoSrc1 and CoSrc2 using synthetic peptides containing substrate motifs for various protein kinases. The specificities of CoSrc1 and CoSrc2 were similar to each other, and to that of mammalian c-Src, in that they showed the strongest phosphorylation of the Src substrate peptide, followed by the Abl substrate. Neither showed any activity toward the protein kinase A substrate, consistent with the prediction that they are tyrosine-specific kinases (Figure 3-3B).

In addition to their involvement in autoinhibitory interactions, the SH2 and SH3 domains of Src family kinases play a role in substrate recognition. This can be recapitulated *in vitro*; mammalian Src family kinases show higher activity toward peptide substrates to which SH3 or SH2 ligands are attached [29, 106]. *M. brevicollis* MbSrc1 displays enhanced phosphorylation of peptides containing SH3 ligands, but not SH2 ligands [93]. We tested *Capsaspora* CoSrc1 and CoSrc2 with substrates possessing SH2 or SH3 ligands (Figure 3-3C). CoSrc1 and CoSrc2 both phosphorylated the substrate sequence to a higher degree when the SH3 ligand was incorporated into the peptide. CoSrc1 showed a 6.5-fold enhancement (102.1 vs 15.5 μmol min^{-1} mg^{-1}), whereas CoSrc2 showed a 7.8-fold enhancement (151.6 vs 19.4 μmol min^{-1} mg^{-1}). Neither enzyme displayed enhanced recognition of a substrate with an SH2 ligand.

Mammalian Src autophosphorylates at the activation loop tyrosine (Tyr-416). This modification increases kinase activity by destabilizing the autoinhibited conformation. To test whether the *Capsaspora* Src kinases are regulated in a similar manner, we first dephosphorylated the purified Src proteins by treating them with YOP tyrosine phosphatase and then incubated them with ATP to promote autophosphorylation. While dephosphorylated CoSrc1 and CoSrc2 had no detectable activity, the ATP-treated enzymes were active (Figure 3-4A). We identified the major site of autophosphorylation in ATP-treated CoSrc2 by liquid chromatography and tandem mass spectrometry of tryptic peptides. As expected, the enzyme is autophosphorylated on Tyr-471, the equivalent of the activation loop Tyr-416 in mammalian Src family kinases (Figure 3-4B).

To examine the activities of CoSrc1 and CoSrc2 in a cellular context, we expressed them in mammalian Src/Yes/Fyn-deficient fibroblasts [107]. These cells lack all Src family kinases and present a low background level of tyrosine phosphorylation. We measured overall tyrosine kinase activity by anti-phosphotyrosine Western blotting. CoSrc2 (or mammalian c-Src) showed robust autophosphorylation and phosphorylation of additional SYF cell proteins, while CoSrc1 showed very little activity (Figure 3-5).

We tested the ability of CoSrc1 and CoSrc2 to functionally substitute for mammalian Src using a Src-responsive reporter gene assay. This assay depends on Src phosphorylation of STAT (signal transducers and activators of transcription) molecules. CoSrc1 and CoSrc2 both

displayed activity in this assay. Compared to mammalian c-Src, CoSrc1 was less active, while CoSrc2 had approximately 2-fold higher activity than c-Src. These results generally match the amount of autophosphorylation seen when expressing CoSrc1 and CoSrc2 in SYF cells (Figure 3-6).

Cloning CoCsk from *C. owczarzaki*. A C-terminal Src kinase (Csk) homologue was identified in the genome of *C. owczarzaki*. We aligned the sequence with the *M. brevicollis* Csk (MbCsk) and mammalian Csk (mCsk). CoCsk contains the conserved SH2, SH3, and kinase domain arrangement (Figure 3-7). A phylogenetic tree comparison is shown in Figure 3-2. CoCsk possesses all six of the residues previously identified as being important in Src recognition (Figure 3-8) [108]. MbCsk is lacking two of these Src recognition sites.

CoCsk Has No Detectable Activity. We cloned and expressed full-length CoCsk in bacterial cells as a fusion protein with glutathione S-transferase (GST). We also produced the corresponding mammalian and *M. brevicollis* GST-Csk constructs. We purified the GST-tagged proteins by glutathione-agarose chromatography and measured activity toward poly(Glu_4Tyr), a general tyrosine kinase substrate (Figure 3-9). In contrast to mammalian Csk or *M. brevicollis* Csk, CoCsk was completely inactive. CoCsk showed no activity toward the Src peptide substrate or any other substrates tested. We tested for phosphorylation of CoSrc1 and CoSrc2 by CoCsk using purified proteins and autoradiography. To reduce the level of background phosphorylation in these experiments, we produced mutant forms of CoSrc1 and CoSrc2 lacking the major sites of autophosphorylation (Y447F and Y471F, respectively). No CoCsk activity toward the *Capsaspora* Src family kinases, the presumed substrates of CoCsk, was detectable (Figure 3-10A,B). Mammalian Csk was also unable to phosphorylate CoSrc1 or CoSrc2 *in vitro*, although phosphorylation of c-Src was detected (Figure 3-10C).

We also tested the activity of CoCsk toward CoSrc1 and CoSrc2 in mammalian cells. Autophosphorylation site mutant forms of CoSrc1 and CoSrc2 were expressed in mammalian SYF cells in the presence and absence of CoCsk. No additional phosphorylation of CoSrc1 and CoSrc2 was observed in cells co-expressing CoCsk (Figure 3-11). To test for binding between CoCsk and CoSrc enzymes, we mixed immobilized GST-CoCsk with lysates from SYF cells

expressing CoSrc1 or CoSrc2. Both forms of CoSrc bound to CoCsk, suggesting the possibility that CoCsk possesses a scaffolding function in *Capsaspora* cells (Figure 3-12A).

Capsaspora CoCsk shows homology with its counterparts from mammals and choanoflagellates (Figure 3-7). Furthermore, the catalytic domain of CoCsk contains all of the conserved residues important for ATP binding and catalysis (Figure 3-8). However, unlike the Src kinase domain, which is constitutively active without the presence of its regulatory domains, the kinase domain of Csk requires interactions with its SH2 and SH3 domains for activity [59]. Mutagenesis experiments highlighted interactions between the β3-αC loop in the kinase domain and multiple amino acids in the SH2 domain that are necessary for Csk activity [109]. The SH3-SH2 linker region is also crucial for Csk activation [110]. In CoCsk, there are approximately 20 amino acids in these regions that differ from the residues conserved in mammalian, fly, sponge, and *M. brevicollis* Csk kinases; these sequence variations could explain the lack of detectable CoCsk activity. We tested this possibility by site-directed mutagenesis of CoCsk Gly-278, which is in the kinase-SH2 interface. In mammalian, fly, sponge, and *M. brevicollis* Csk kinases, an asparagine residue is present at this position. An alanine substitution at this position in mammalian Csk was previously found to decrease activity [110]. We produced a G278N mutant form of CoCsk, expressed and purified the GST fusion protein, and tested its activity against poly(Glu_4Tyr). The mutation led to a partial recovery of CoCsk activity (specific activity of 0.5 μmol min^{-1} mg^{-1}) (Figure 3-12B). These results suggest that Gly-278 and other residues in the interface contribute to the low activity of CoCsk. Importantly, these results also argue that the low activity of CoCsk is an intrinsic feature of the protein, rather than (for example) an artifact of protein expression in *E. coli.*

CoCsk Does Not Inhibit CoSrc. Phosphorylation of the C-termini of mammalian Src family kinases by Csk promotes the closed, inactive form. Although we were not able to detect any phosphorylation of CoSrc1 or CoSrc2 by CoCsk *in vitro* or in SYF cells, we tested for any inhibitory effect of CoCsk. No significant change in *in vitro* activity of purified CoSrc1 and CoSrc2 was seen upon incubation with CoCsk (Figure 3-13A). Next, we expressed CoSrc1 and CoSrc2 in SYF cells in the presence or absence of CoCsk. We immunoprecipitated CoSrc1 and CoSrc2 and measured their activities toward a synthetic Src substrate. In contrast to the

inhibitory effect of Csk on mammalian Src kinases, co-expression of CoCsk led to modest increases in CoSrc activity (Figure 3-13B). We also tested CoSrc variants in which the C-terminal tyrosines had been mutated to phenylalanine, eliminating the possibility of negative regulation (CoSrc1 Y558F and CoSrc2 Y582F). The mutant forms of CoSrc1 and CoSrc2 were expressed in SYF cells, immunoprecipitated, and compared to wild-type CoSrc1 and CoSrc2 co-expressed with CoCsk. We observed no significant difference in CoSrc activity between the mutant (and presumably maximally active) forms and wild-type CoSrc isolated from CoCsk-expressing cells (Figure 3-13C). These results suggest that CoCsk is incapable of promoting CoSrc inhibition. Consistent with these findings, anti-phosphotyrosine Western blotting of SYF cell lysates showed comparable levels of overall cellular phosphorylation in the presence or absence of CoCsk (Figure 3-14).

To study endogenous CoSrc1 and CoSrc2 in *C. owczarzaki* cells, we generated specific antipeptide antibodies and carried out immunofluorescence experiments on fixed *Capsaspora owczarzaki* cells. These localization assays showed that CoSrc1 and CoSrc2 had similar localization patterns, with both clustering at specific points along filopodia as well as in the cell body. These points of localization matched those seen along filopodia when cells were probed with an actin antibody (Figure 3-15). We used the CoSrc1- and CoSrc2-specific antibodies to immunoprecipitate the enzymes from *Capsaspora* lysates. Endogenous CoSrc1 and CoSrc2 isolated from *Capsaspora* cells were active toward a synthetic Src substrate peptide (Figure 3-16A,B). The levels of activity (approximately 100-150 pmol of phosphate/1.5 mg of lysate) were on the same order of magnitude as that of CoSrc1 and CoSrc2 overexpressed in SYF cells (approximately 20 or 500 pmol of phosphate/mg, respectively) (Figure 3-16A,B). We attempted to analyze the amounts of immunoprecipitated CoSrc by Western blotting. Overexpressed CoSrc1 and CoSrc2 in SYF cells could easily be visualized, while levels of endogenous CoSrc1 and CoSrc2 protein in *Capsaspora* cells were undetectable (Figure 3-16C). These experiments suggest that the specific activities of CoSrc1 and CoSrc2 in *Capsaspora* cells are higher than those measured in mammalian cells, consistent with a lack of Csk-mediated regulation.

Discussion

Functional studies of the ancestral counterparts to mammalian signaling molecules can provide a window into their evolution. Choanoflagellates, the closest known living relatives of metazoans, possess numerous tyrosine kinases. The Src and Csk family kinases from the choanoflagellates *M. brevicollis* and *M. ovata* are broadly similar to the mammalian families in their domain architectures and enzymatic function. In both species, Csk phosphorylates Src but does not tightly regulate activity [87, 93]. The *M. brevicollis* enzymes showed some additional differences from mammalian c-Src, including a lack of SH2-directed substrate recognition and a novel C2 lipid-binding domain in one isoform (MbSrc4) [93, 94]. Because these signaling molecules most likely emerged before the split between metazoan and choanoflagellates, we have analyzed the genome of their sister lineage, *C. owczarzaki*. We have investigated the properties of its two Src-like kinases and one Csk, comparing them to their mammalian counterparts to discover their functional similarities and differences.

The specific activities of CoSrc1 and CoSrc2 toward a synthetic peptide containing an optimal Src recognition sequence are comparable to that of mammalian c-Src. The substrate specificities of CoSrc1 and CoSrc2 also resemble that of c-Src; both phosphorylated the Src substrate synthetic peptide with a greater efficiency than peptides with recognition sequences for other tyrosine kinases substrates (Figure 3-3B). CoSrc1 and CoSrc2 also displayed the same dependence on the phosphorylation of the activation loop tyrosine observed for mammalian c-Src (Figure 3-4). In contrast, although CoSrc1 and CoSrc2 both showed evidence of SH3 substrate targeting as measured by synthetic peptides, the SH2 domain was not coupled to substrate recognition (Figure 3-3C). These results are similar to those seen for the *M. brevicollis* MbSrc1 [93]. The absence of SH2 targeting in both *Monosiga* and *Capsaspora* Src kinases suggests that functional coupling of the SH2 and kinase domains and autoregulation did not develop until later in SFK evolution. This lower degree of autoregulation may explain the results of the Src reporter assay (Figure 3-6). The high activity of CoSrc2 as compared to c-Src may be a result of higher basal activity due to an undeveloped autoregulatory system in these unicellular organisms.

Genomic analysis has shown that *C. owczarzaki* contains almost all of the necessary components of mammalian focal adhesions, although these complexes were previously thought to exist only in metazoans. Indeed, *Capsaspora* is the only nonmetazoan group that possesses an integrin-mediated adhesion pathway with all the same components present in the integrin machinery of metazoans. Choanoflagellates have secondarily lost the integrin adhesion pathway and contain only some of the signaling and scaffolding proteins associated with the integrin adhesion apparatus such as Src and talin [105]. This shows that the origins of the protein precursors of the mammalian integrin adhesion complexes predate the transition to multicellularity in the metazoan lineage. Localization experiments showed a general pattern of the presence of CoSrc1 and CoSrc2 in the filopodia of *Capsaspora* cells, co-localizing with actin (Figure 3-15). *Capsaspora* cells use these filopodia to crawl along a substrate. Thus, a potential explanation is that they use focal adhesion-like complexes to generate force for moving. We hypothesize that ancestral metazoans co-opted this nearly intact machinery to create the modern focal adhesion system. Unfortunately, it is not at present possible to test the physiological roles of CoSrc and CoCsk in intact *Capsaspora* cells, because the technology does not yet exist to introduce genes or to block their function in this organism.

Functional studies of the *Capsaspora* Csk revealed an unexpected difference between this protein and all previously studied Csk kinases, including those from the choanoflagellates, *M. ovata*, and *M. brevicollis*. In mammalian cells, Csk plays a crucial role in regulating Src activity by phosphorylating a conserved tyrosine at the C-terminal tail. We found that CoCsk displayed no measurable activity toward a synthetic kinase substrate or toward CoSrc (Figures 3-9, 3-10). CoCsk was expressed and purified using exactly the same methodology that was used for mammalian Csk [59] and *M brevicollis* Csk [93]. Furthermore, a G278N mutation resulted in a partial recovery of CoCsk activity (Figure 3-12B), demonstrating that the experimental procedures used to express and assay CoCsk were compatible with the measurement of CoCsk activity. Using our protocol with [γ-^{32}P]ATP and poly(Glu_4-Tyr) as substrates, we conservatively estimate the lower detection limit of our assay with wild-type CoCsk to be a specific activity of 2.5 nmol min^{-1} mg^{-1}, which is approximately 50 times lower than the activity seen for mammalian Csk (121 nmol min^{-1} mg^{-1}. CoCsk also had no inhibitory effect on CoSrc activity in vitro or in cells (Figure 3-12). In the immunoprecipitation-kinase assays from SYF

cells (Figure 3-13B), the presence of CoCsk actually enhanced CoSrc activity, perhaps because of the formation of a complex between the two proteins (Figure 3-12A); such a complex could stabilize the open form of CoSrc. Immunoprecipitation-kinase assays from *C. owczarzaki* cells showed that CoSrc1 and CoSrc2 were undetectable in *Capsaspora* lysates, suggesting that the specific activities are high. This is consistent with a lack of endogenous CoCsk activity. A high basal level of Src activity was also observed in *M. ovata* [87]. Csk may have evolved as an inactive kinase or alternatively as a kinase that did not phosphorylate Src; in the latter case, Csk could have lots its activity in the filasterean lineage. Our work on *M. vibrans*, the other known member of the filasterean lineage, has shown that the lack of CoCsk activity is not an anomaly. MvCsk lacks detectable enzymatic activity as well, indicating that the lack of Src down-regulation is a hallmark of the filasterean clade (Chapter 4). In metazoan cells, Csk is essential for preventing unbridled Src activity; it is likely that the development of the intimate Csk-Src relationship was crucial in the evolution of signaling systems in multicellular animals.

Figure 3-1. **Amino acid sequence alignment of *C. owczarzaki* CoSrc1 and CoSrc2, *M. brevicollis* MbSrc1, and human c-Src**. The SH3 domain is boxed in red, the SH2 domain boxed in blue, and the kinase domain boxed in green. The orange circles represent conserved regulatory tyrosines: Tyr-416 in the activation loop and Tyr-527 in the C-terminal tail.

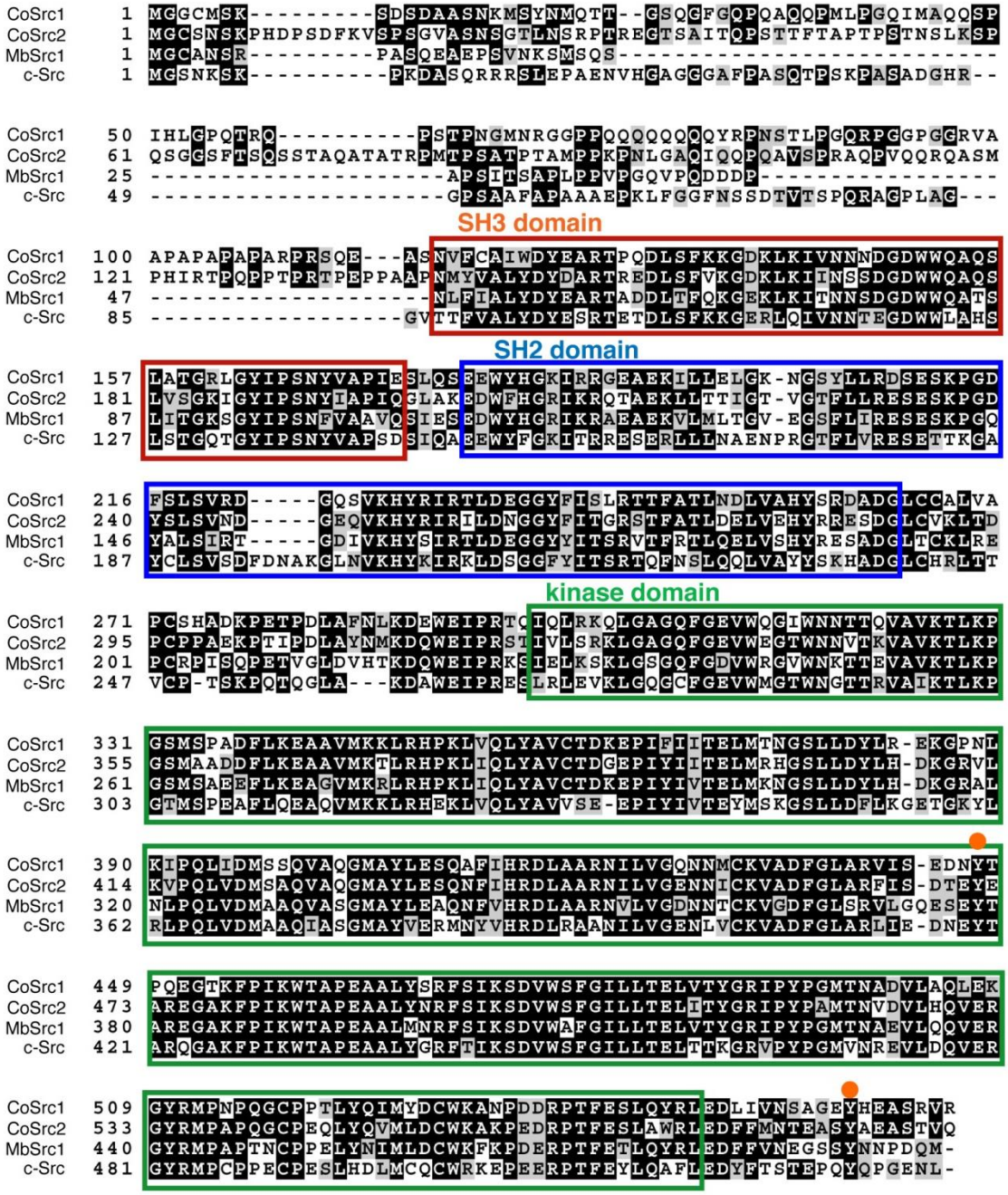

SH3 domain
SH2 domain
kinase domain

Figure 3-2**. Phylogenetic tree of four cytoplasmic TK families**. An alignment of 363 amino acid sites was used for the tree inference. The alignment covers SH3, SH2, and TK domains. GenBank accession numbers or Flybase IDs of sequences are given in brackets. Data for *A. queenslandica, M. brevicollis,* and *C. owczarzaki* were from the whole genome sequence. The numbers on branches indicate the statistical nodal support as obtained from 100 maximum likelihood bootstrap replicates using the WAG model of evolution.

Figure made by Dr. Hiroshi Suga, University of Barcelona

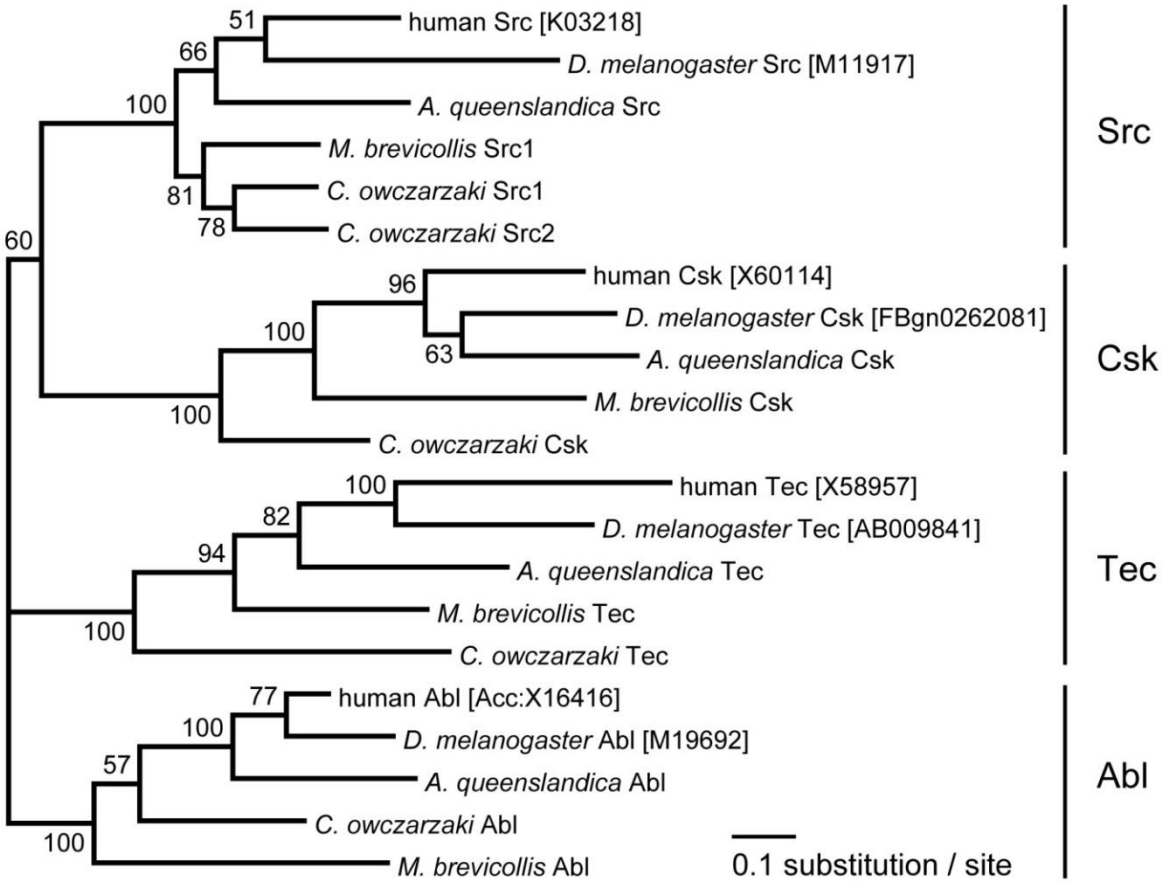
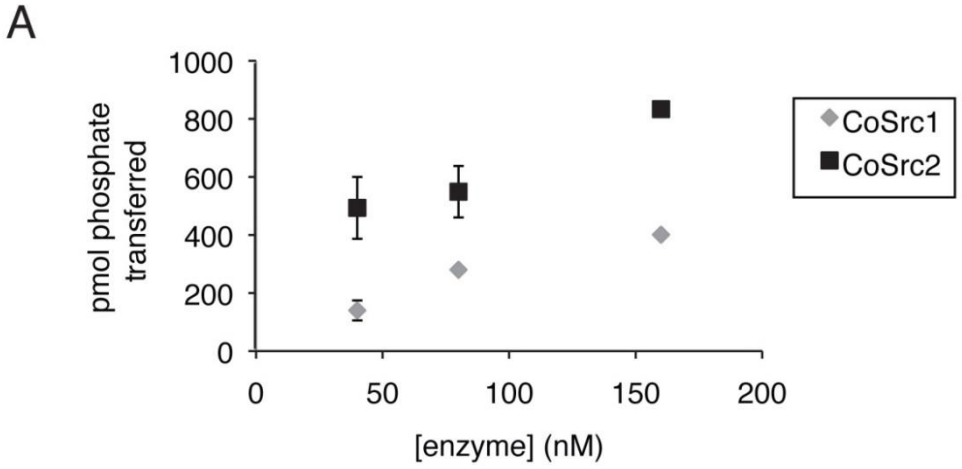
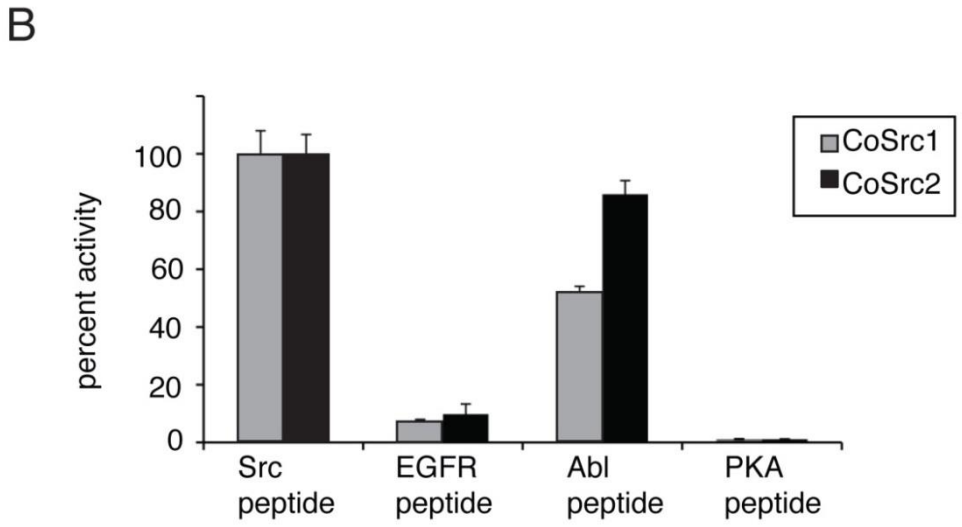

human Src [K03218]
D. melanogaster Src [M11917]
A. queenslandica Src
M. brevicollis Src1
C. owczarzaki Src1
C. owczarzaki Src2
human Csk [X60114]
D. melanogaster Csk [FBgn0262081]
A. queenslandica Csk
M. brevicollis Csk
C. owczarzaki Csk
human Tec [X58957]
D. melanogaster Tec [AB009841]
A. queenslandica Tec
M. brevicollis Tec
C. owczarzaki Tec
human Abl [Acc:X16416]
D. melanogaster Abl [M19692]
A. queenslandica Abl
C. owczarzaki Abl
M. brevicollis Abl
0.1 substitution / site
Src
Csk
Tec
Abl

Figure 3-3. **CoSrc1 and CoSrc2 are active tyrosine kinases.** (A) Activity of CoSrc1 and CoSrc2 at varying enzyme concentrations. Enzyme activities were measured with 0.5 mM Src peptide substrate and 0.25 mM ATP using the phosphocellulose paper binding assay. Reactions proceeded for 5 min at 30°C. (B) Substrate specificities of CoSrc2 and CoSrc2 were investigated using synthetic peptides incorporating recognition motifs from four protein kinases. Enzymes were incubated with 0.25 mM [γ-^{32}P]ATP and 200 µM peptide, and reactions proceeded for 20 min at 30°C. Activities are presented relative to Src peptide (100%) for each enzyme. (C) Substrate targeting by CoSrc1 and CoSrc2. Enzymes (400 nM) were tested with synthetic peptide substrates containing SH3 or SH2 ligand sequences or matched controls (see Materials and Methods for peptide sequences). Reactions proceeded for 30 min at 30°C, and mixtures were analyzed by the phosphocellulose paper assay. All assays were performed in duplicate, and error bars show standard deviations.


A
1000
800
600
400
200
0
pmol phosphate transferred
0
50
100
150
200
[enzyme] (nM)
CoSrc1
CoSrc2


B
100
80
60
40
20
0
percent activity
Src peptide
EGFR peptide
Abl peptide
PKA peptide
CoSrc1
CoSrc2

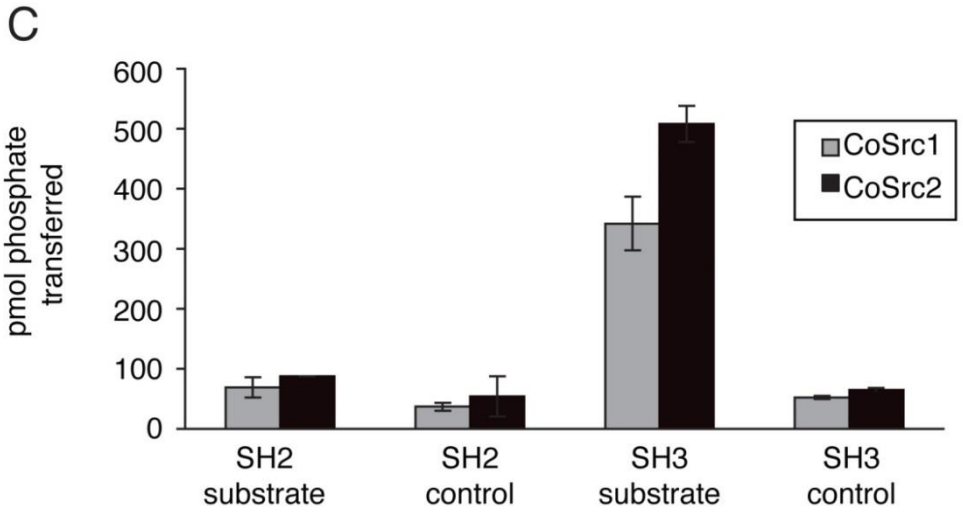
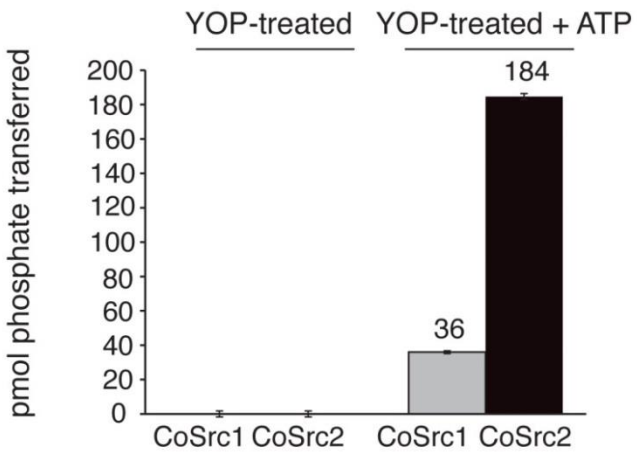

C
600
500
400
300
200
100
0
pmol phosphate transferred
SH2 substrate
SH2 control
SH3 substrate
SH3 control
CoSrc1
CoSrc2

Figure 3-4. **Effect of autophosphorylation on CoSrc activity.** (A) CoSrc1 and CoSrc2 were treated with immobilized GST-YOP for 30 min at 30°C. The YOP was removed by centrifugation, and the mixture was assayed directly (left bars) or after incubation with ATP for 30 min at 30°C (right bars). Activity measurements were performed in duplicate with the Src synthetic peptide and the phosphocellulose paper assay, as in Figure 3-2. (B) . Analysis of the major autophosphorylation site in CoSrc2 by mass spectrometry. YOP-treated CoSrc2 was incubated in the presence of ATP for 20 min at 30° C, and digested overnight with trypsin. A control sample of YOP-treated CoSrc2 was analyzed without ATP treatment. The samples were analyzed by liquid chromatography-tandem mass spectrometry on a Thermo-Fisher Scientific LTQ XL instrument. The liquid chromatography elution profile of the activation loop peptide containing Y471 (phosphorylated or unphosphorylated) is shown in the top panel. MS results (bottom) show the characteristic +80 shift for the phosphorylated Y471 peptide.

A


YOP-treated
YOP-treated + ATP
pmol phosphate transferred
200
180
160
140
120
100
80
60
40
20
0
184
36
CoSrc1 CoSrc2
CoSrc1 CoSrc2

B

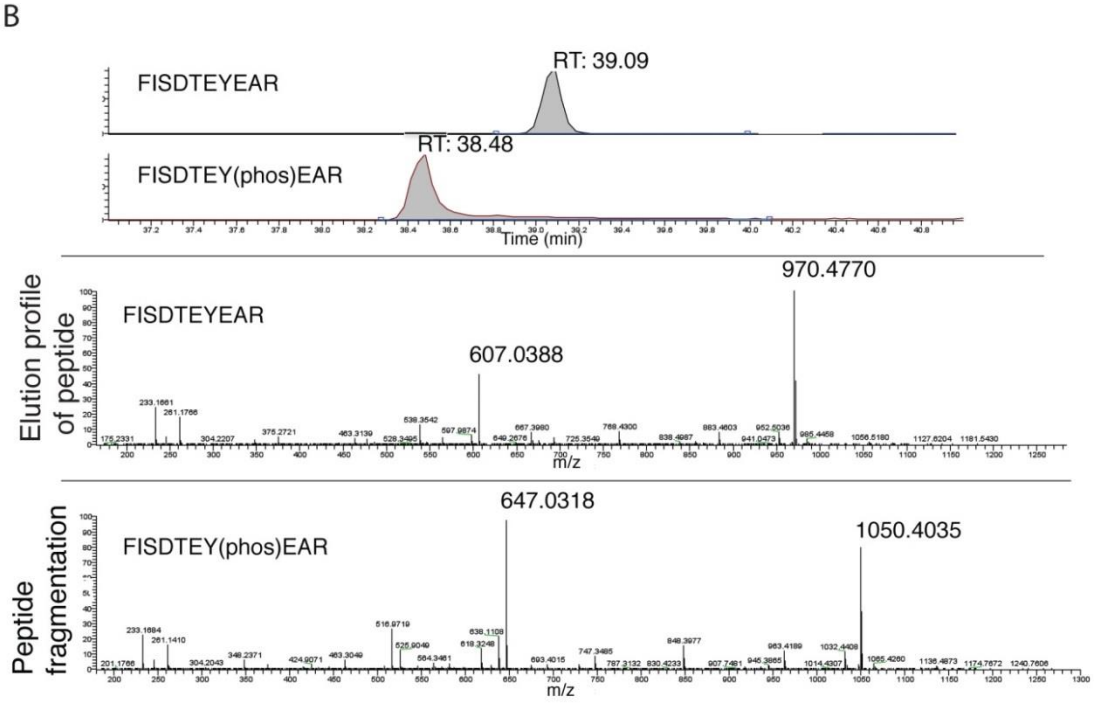

FISDTEYEAR
RT: 39.09
RT: 38.48
FISDTEY(phos)EAR
Time (min)
Elution profile of peptide
FISDTEYEAR
970.4770
607.0388
m/z
Peptide fragmentation
FISDTEY(phos)EAR
647.0318
1050.4035
m/z

Figure 3-5. **CoSrc autophosphorylation in SYF cells.** Whole cell lysates from untransfected SYF cells, or SYF cells expressing mammalian c-Src, CoSrc1, or CoSrc2, were analyzed by Western blotting with anti-phosphotyrosine antibody. The membrane was stripped and re-probed with anti-FLAG antibody to compare Src expression. The blots were quantitated by dividing the signal for autophosphorylated Src with the signal for total Src (ImageJ), and the value for c-Src was set to 1.0.

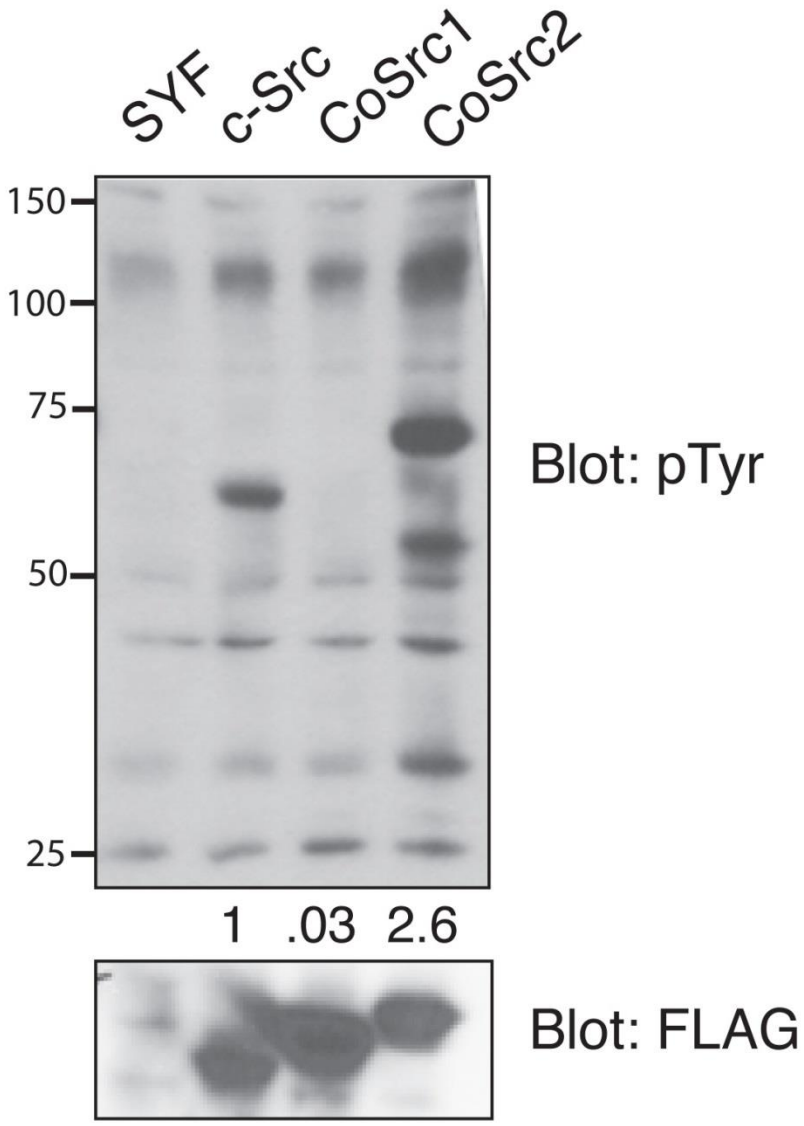

SYF
c-Src
CoSrc1
CoSrc2
150
100
75
50
25
Blot: pTyr
1
.03
2.6
Blot: FLAG

Figure 3-6. **CoSrc1 and CoSrc2 can functionally replace mammalian Src.** Luciferase activity was measured in lysates from SYF cells transfected with the indicated plasmids and the GAS luciferase reporter plasmid. Experiments were performed in triplicate, and error bars show S.D.

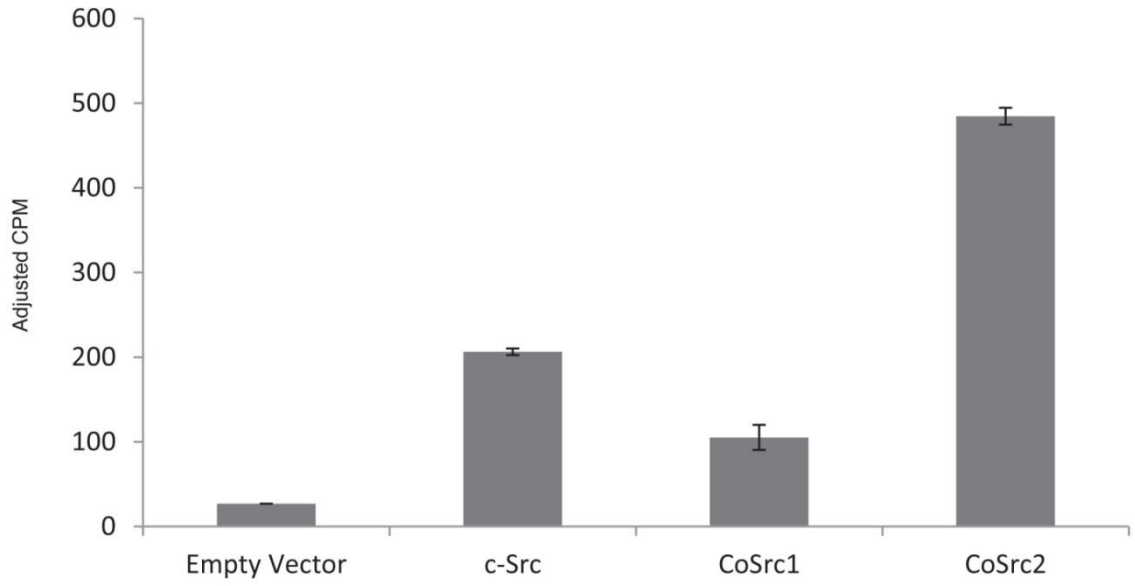

600
500
400
300
200
100
0
Adjusted CPM
Empty Vector
c-Src
CoSrc1
CoSrc2

Figure 3-7. **Amino acid sequence alignment of *C. owczarzaki* CoCsk, *M. brevicollis* MbCsk, and human Csk.** The SH3 domain is boxed in red, the SH2 domain boxed in blue, and the kinase domain boxed in green. The orange circles represent sites that are important for interaction with c-Src.

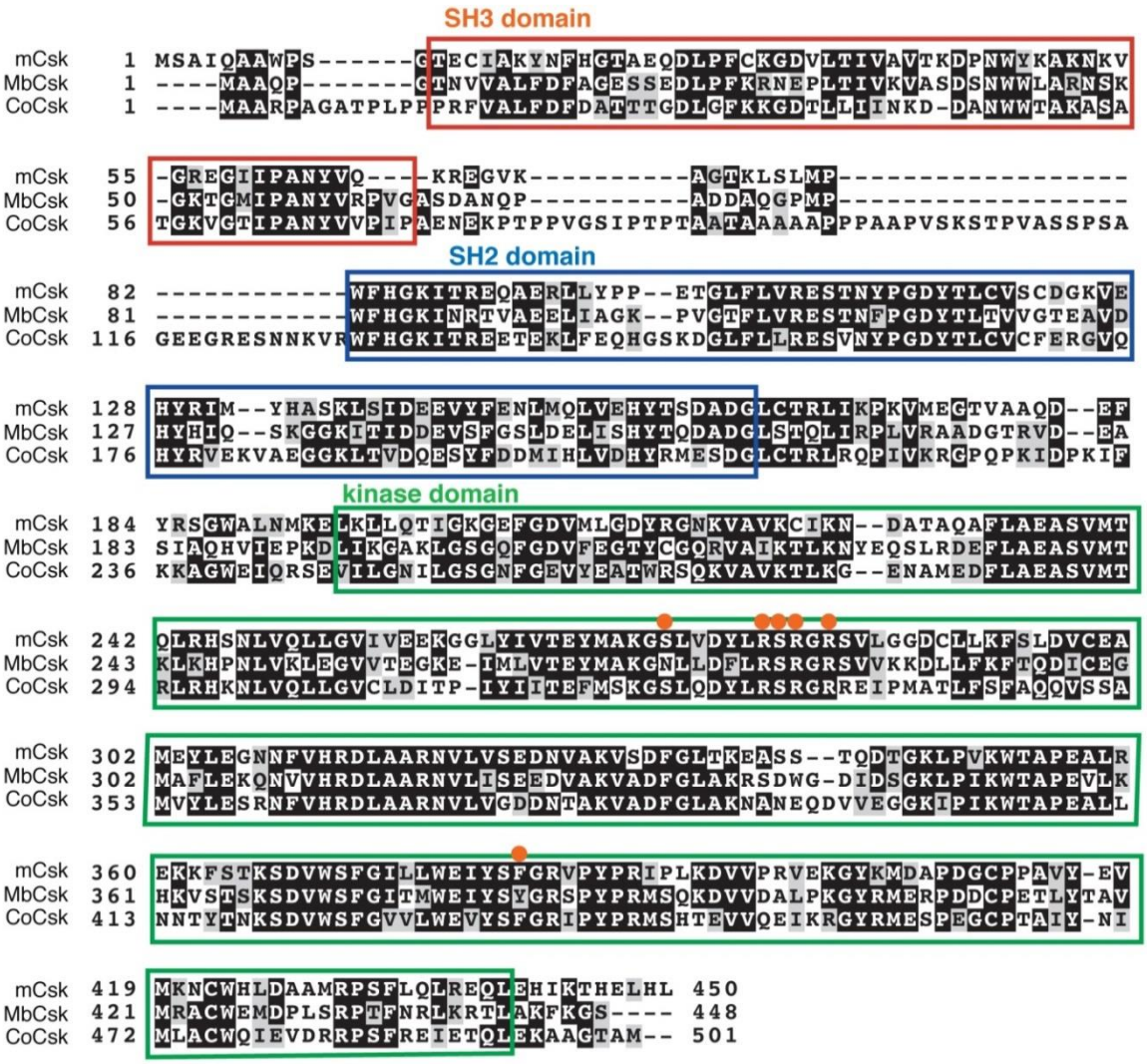

SH3 domain
SH2 domain
kinase domain
mCsk 1 MSAIQAAWPS------GTECIAKYNFHGTAEQDLPFCKGDVLTIVAVTKDPNWYKAKNKV
MbCsk 1 ----MAAQP-------GTNVVALFDFAGESSEDLPFKRNEPLTIVKVASDSNWWLARNSK
CoCsk 1 ----MAARPAGATPLPPPRFVALFDFDATTTGDLGFKKGDTLLIINKD-DANWWTAKASA
mCsk 55 -GREGIIPANYVQ----KREGVK----------AGTKLSLMP----------
MbCsk 50 -GKTGMIPANYVRPVGASDANQP----------ADDAQGPMP----------
CoCsk 56 TGKVGTIPANYVVPIPAENEKPTPPVGSIPTPTAATAAAAAPPPAAPVSKSTPVASSPSA
mCsk 82 ----------WFHGKITREQAERLLYPP--ETGLFLVRESTNYPGDYTLCVSCDGKVE
MbCsk 81 ----------WFHGKINRTVAEELIAGK--PVGTFLVRESTNFPGDYTLTVVGTEAVD
CoCsk 116 GEEGRESNNKVRWFHGKITREETEKLFEQHGSKDGLFLLRESVNYPGDYTLCVCFERGVQ
mCsk 128 HYRIM--YHASKLSIDEEVYFENLMQLVEHYTSDADGLCTRLIKPKVMEGTVAAQD--EF
MbCsk 127 HYHIQ--SKGGKITIDDEVSFGSLDELISHYTQDADGLSTQLIRPLVRAADGTRVD--EA
CoCsk 176 HYRVEKVAEGGKLTVDQESYFDDMIHLVDHYRMESDGLCTRLRQPIVKRGPQPKIDPKIF
mCsk 184 YRSGWALNMKELKLLQTIGKGEFGDVMLGDYRGNKVAVKCIKN--DATAQAFLAEASVMT
MbCsk 183 SIAQHVIEPKDLIKGAKLGSGQFGDVFEGTYCGQRVAIKTLKNYEQSLRDEFLAEASVMT
CoCsk 236 KKAGWEIQRSEVILGNILGSGNFGEVYEATWRSQKVAVKTLKG--ENAMEDFLAEASVMT
mCsk 242 QLRHSNLVQLLGVIVEEKGGLYIVTEYMAKGSLVDYLRSRGRSVLGGDCLLKFSLDVCEA
MbCsk 243 KLKHPNLVKLEGVVTEGKE-IMLVTEYMAKGNLLDFLRSRGRSVVKKDLLFKFTQDICEG
CoCsk 294 RLRHKNLVQLLGVCLDITP-IYIITEFMSKGSLQDYLRSRGREIPMATLFSFAQQVSSA
mCsk 302 MEYLEGNNFVHRDLAARNVLVSEDNVAKVSDFGLTKEASS--TQDTGKLPVKWTAPEALR
MbCsk 302 MAFLEKQNVVHRDLAARNVLISEEDVAKVADFGLAKRSDWG-DIDSGKLPIKWTAPEVLK
CoCsk 353 MVYLESRNFVHRDLAARNVLVGDDNTAKVADFGLAKNANEQDVVEGGKIPIKWTAPEALL
mCsk 360 EKKFSTKSDVWSFGILLWEIYSFGRVPYPRIPLKDVVPRVEKGYKMDAPDGCPPAVY-EV
MbCsk 361 HKVSTSKSDVWSFGITMWEIYSYGRSPYPRMSQKDVVDALPKGYRMERPDDCPETLYTAV
CoCsk 413 NNTYTNKSDVWSFGVVLWEVYSFGRIPYPRMSHTEVVQEIKRGYRMESPEGCPTAIY-NI
mCsk 419 MKNCWHLDAAMRPSFLQLREQLEHIKTHELHL 450
MbCsk 421 MRACWEMDPLSRPTFNRLKRTLAKFKGS---- 448
CoCsk 472 MLACWQIEVDRRPSFREIETQLEKAAGTAM-- 501

Figure 3-8. **CoCsk contains the essential tyrosine kinase catalytic residues**. (A) Alignment of the kinase domains of mammalian Csk (mCsk), *Monosiga brevicollis* (MbCsk) and CoCsk. The conserved catalytic loop residues (signified by an orange box) and the kinase-conserved DFG residues in the activation loop (shown as the purple box) are both present. (B) Structure of the CoCsk catalytic domain, generated by modeling onto the structure of the mCsk catalytic domain (pdb code = 1byg) using the HHpred toolkit (http://toolkit.tuebingen.mpg.de/hhpred). The catalytic loop and DFG residues are colored as in (A).

A

```
mCsk   188 WALNMKELKLLQTIGKGEFGDVMLGDYRGNKVAVKCIKN--DATAQAFLAEASVMTQLRH
MbCsk  187 HVIEPKDLIKGAKLGSGQFGDVFEGTYCGQRVAIKTLKNYEQSLRDEFLAEASVMTKLKH
CoCsk  241 WEIQRSEVILGNILGSGNFGEVYEATWRSQKVAVKTLKG--ENAMEDFLAEASVMTRLRH

mCsk   246 SNLVQLLGVIVEEKGGLYIVTEYMAKGSLVDYLRSRGRSVLGGDCLLKFSLDVCEAMEYL
MbCsk  247 PNLVKLEGVVTEGKE-IMLVTEYMAKGNLLDFLRSRGRSVVKKDLLFKFTQDICEGMAFL
CoCsk  299 KNLVQLLGVCLDITP-IYIITEFMSKGSLQDYLRSRGRREIPMATLFSFAQQVSSAMVYL

mCsk   306 EGNNFVHRDLAARNVLVSEDNVAKVSDFGLTKEASSTQDTGKLPVKWTAPEALREKKF
MbCsk  306 EKQNVVHRDLAARNVLISEEDVAKVADFGLAKRSDWGDIDSGKLPIKWTAPEVLKHKVS
CoCsk  358 ESRNFVHRDLAARNVLVGDDNTAKVADFGLAKNANEQDVVEGGKIPIKWTAPEALLNNTY

mCsk   364 STKSDVWSFGILLWEIYSFGRVPYPRIPLKDVVPRVEKGYKMDAPDGCPPAVYEVMKNC
MbCsk  365 TSKSDVWSFGITMWEIYSYGRSPYPRMSQKDVVDALPKGYRMERPDDCPETLYTAVMRAC
CoCsk  418 TNKSDVWSFGVVLWEVYSFGRIPYPRMSHTEVVQEIKRGYRMESPEGCPTAIYNIMLAC

mCsk   423 WHLDAAMRPSFLQLREQLEHIKTHELHL
MbCsk  425 WEMDPLSRPTFNRLKRTLAKFKGS----
CoCsk  477 WQIEVDRRPSFREIETQLEKAAGTAM--
```

B

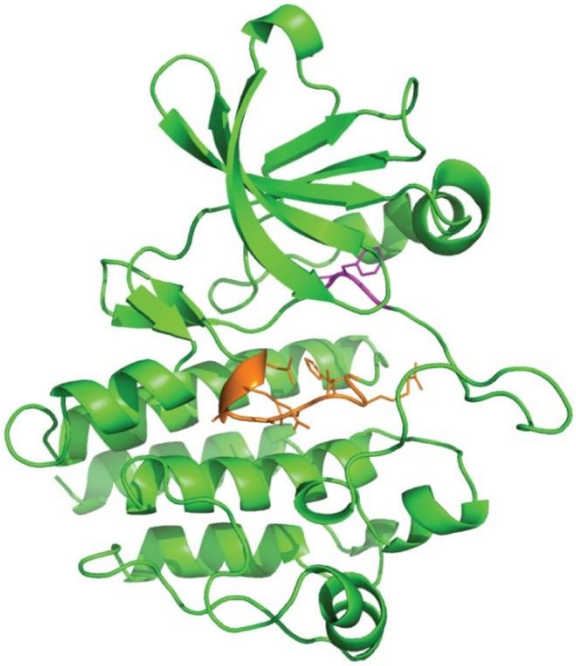


Figure 3-9. **Activity of CoCsk toward a synthetic peptide is undetectable.** Purified CoCsk, *M. brevicollis* Csk (MbCsk), and human Csk (200 nM) were assayed with poly(Glu_4Tyr) (1 mg/mL). Reaction mixtures contained 500 μM [γ-^{32}P]ATP. Activity was analyzed at various time points by scintillation counting.

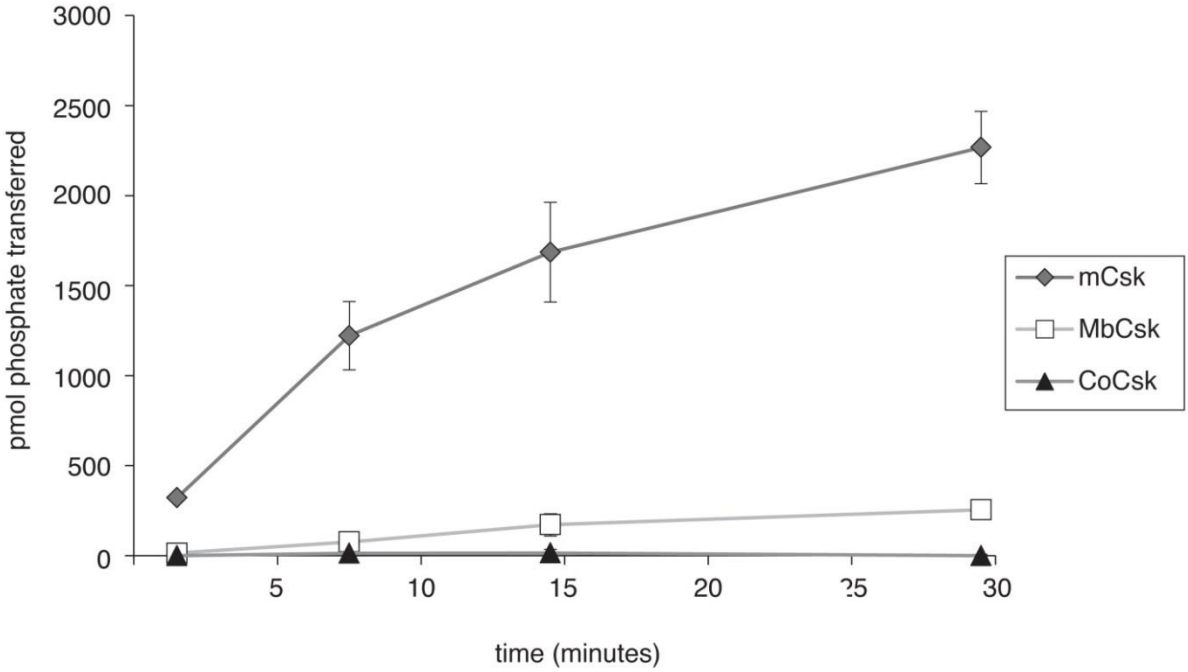
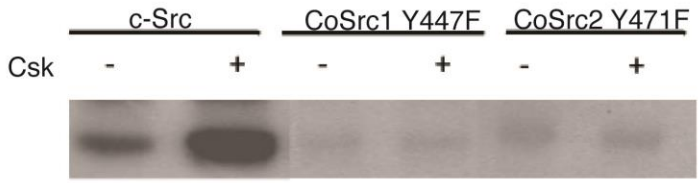
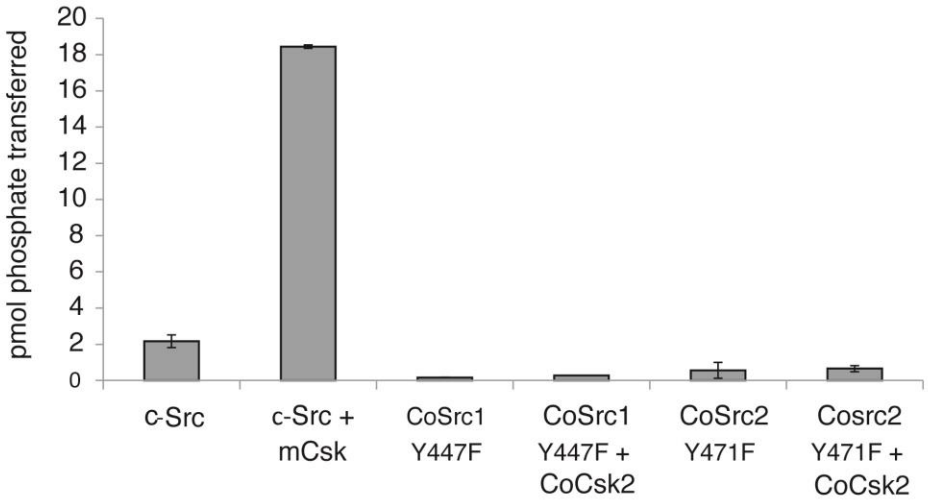

3000
2500
2000
1500
1000
500
0
pmol phosphate transferred
5
10
15
20
25
30
time (minutes)
mCsk
MbCsk
CoCsk

Figure 3-10. **CoCsk has no activity toward CoSrc in vitro.** (A) CoCsk does not phosphorylate CoSrc in vitro. Purified CoCsk (400 nM) was incubated with purified CoSrc1 or CoSrc2 (activation loop mutants, 800 nM) in the presence of [γ-^{32}P]ATP. Reactions were stopped by the addition of Laemmli buffer, and mixtures were analyzed by SDS-PAGE and autoradiography. (B) The bands shown in panel A were excised from the gel, dissolved, and analyzed by scintillation counting. (C) Mammalian Csk does not phosphorylate CoSrc1 or CoSrc2. Kinase-inactive mutants of CoSrc1, CoSrc2, or c-Src catalytic domain (10 μg) were incubated with mammalian Csk (1.8 μg) in the presence of [γ-^{32}P]ATP (0.4 mM for 20 min at 30°C). The reactions were analyzed by SDS-PAGE and autoradiography. Black arrow: position of CoSrc1, open arrow: position of CoSrc2.

A.


c-Src
CoSrc1 Y447F
CoSrc2 Y471F
Csk
-
+
-
+
-
+

B.


pmol phosphate transferred
20
18
16
14
12
10
8
6
4
2
0
c-Src
c-Src + mCsk
CoSrc1 Y447F
CoSrc1 Y447F + CoCsk2
CoSrc2 Y471F
Cosrc2 Y471F + CoCsk2

C.

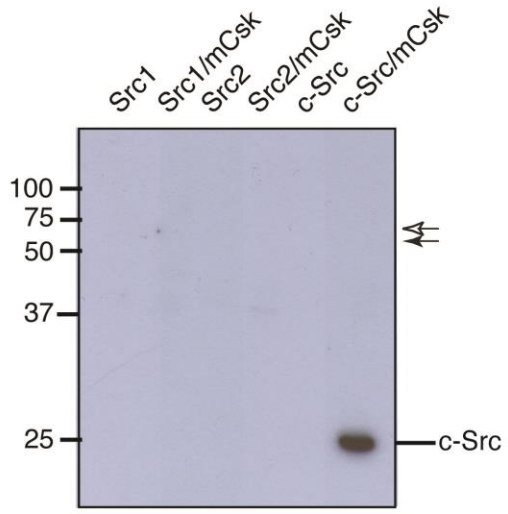

Src1
Src1/mCsk
Src2
Src2/mCsk
c-Src
c-Src/mCsk
100
75
50
37
25
c-Src

Figure 3-11. **CoCsk does not phosphorylate CoSrc in mammalian cells.** Whole cell lysates from SYF cells co-expressing the activation loop mutants (CoSrc1 Y447F or CoSrc2 Y471F) and CoCsk were analyzed by Western blotting with the anti-pTyr antibody. The filled arrow denotes the position of CoSrc1, and the empty arrow denotes the position of CoSrc2. The membrane was stripped and re-probed with anti-FLAG antibody to measure Src expression. The blots were quantitated by dividing the signal for autophosphorylated Src by the signal for total Src. Values in the lanes with CoCsk were compared to the values for CoSrc1 and CoSrc2 alone, which were each set to 1.0.

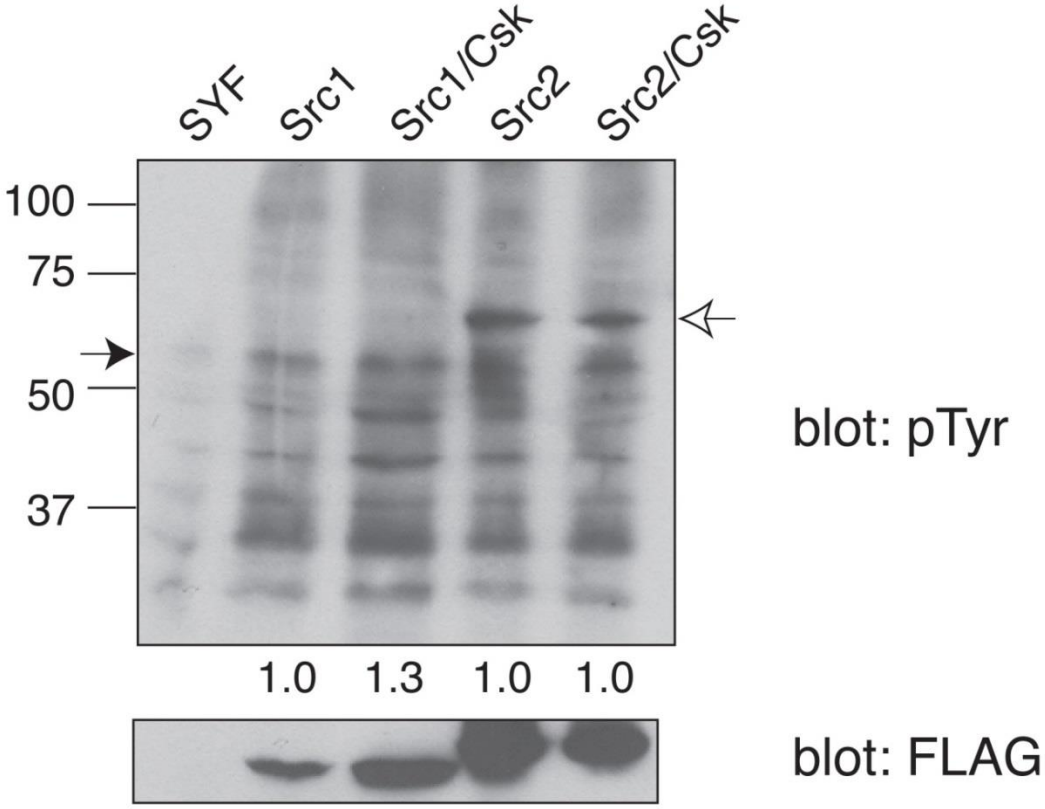

SYF
Src1
Src1/Csk
Src2
Src2/Csk
100
75
50
37
1.0
1.3
1.0
1.0
blot: pTyr
blot: FLAG

Figure 3-12. **Biochemical studies of CoCsk.** (A) Binding of CoSrc1 and CoSrc2 to CoCsk. GST-CoCsk (or GST control) was immobilized on glutathione-agarose beads. Lysates (0.5 mg) From SYF cells expressing CoSrc1 or CoSrc2 were incubated with the resins for 1 h at 4°C. After being washed, bound proteins were analyzed by anti-FLAG Western blotting. For Comparison, 5% of the lysates were analyzed by anti-FLAG Western blotting (bottom). (B) The G278N mutation partially restores CoCsk activity. Purified wild-type GST-CoCsk and GST-CoCsk-G278N were assayed with poly(Glu_4Tyr) (1 mg/mL). Reaction mixtures contained 500 μM [γ-^{32}P]ATP. Activity was analyzed at various time points by scintillation counting.

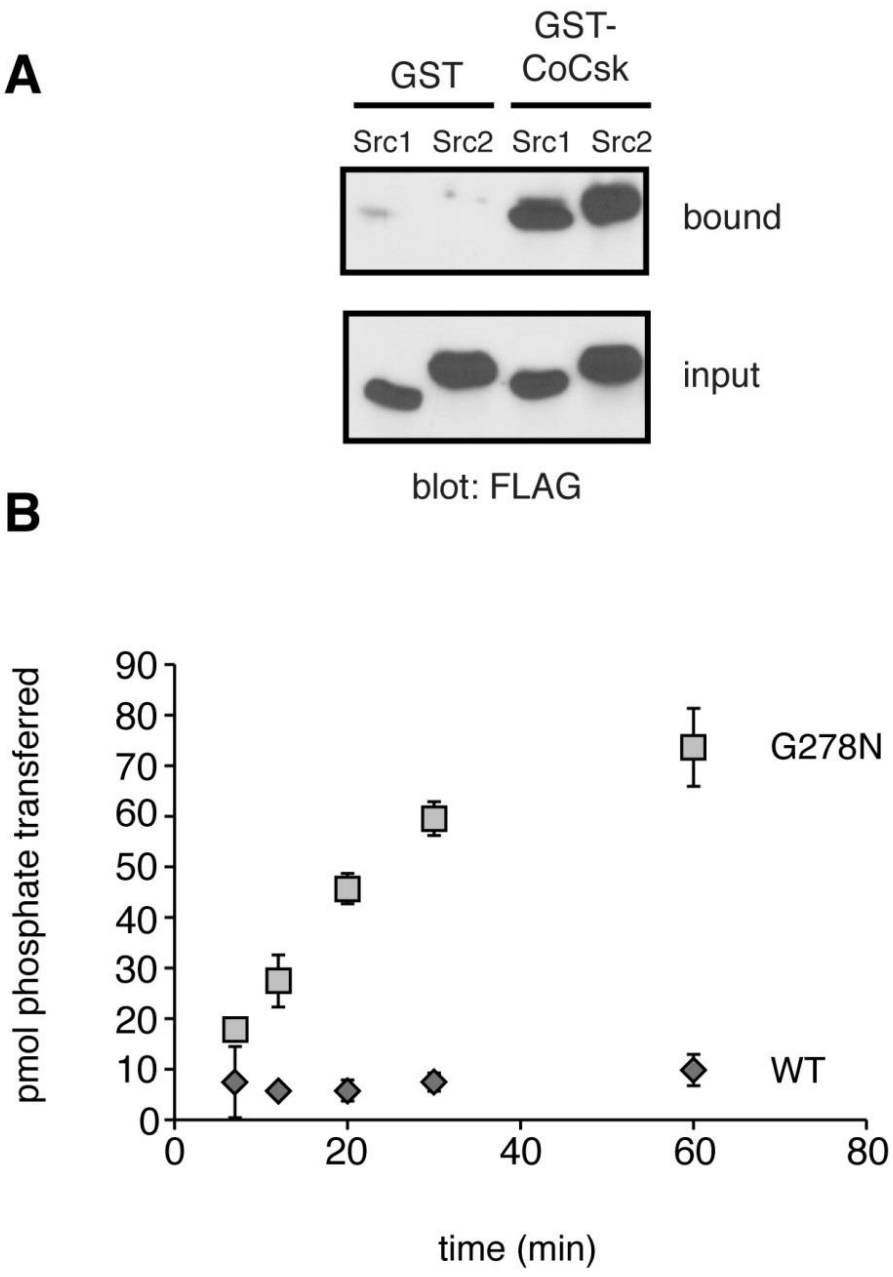

A
GST
GST-
CoCsk
Src1
Src2
Src1
Src2
bound
input
blot: FLAG
B
pmol phosphate transferred
90
80
70
60
50
40
30
20
10
0
0
20
40
60
80
time (min)
G278N
WT

Figure 3-13. **CoCsk does not inhibit CoSrc activity.** (A) In vitro CoSrc kinase assays were performed in the presence or absence of purified CoCsk. Reaction mixtures contained 500 nM CoSrc1 or CoSrc2, CoCsk (600 nM0, 500 μM [γ-^{32}P]ATP and RCM-lysozyme (0.4 mg/mL) as a Src substrate. After 20 min at 30°C, reaction mixtures were analyzed by scintillation counting, as described in Materials and Methods. For each enzyme, activity without CoCsk was normalized to 100%. (B) CoSrc1 or CoSrc2 was expressed in SYF cells in the presence and absence of CoCsk. CoSrc proteins were immunoprecipitated with immobilized FLAG antibody and incubated with Src substrate peptide and [γ-^{32}P]ATP. After 20 min at 30°C, activity was measured using the phosphocellulose paper assay. (C) Similar experiments were conducted to compare wild-type CoSrc1 and CoSrc2 (isolated from SYF cells co-expressing CoCsk) with mutant forms lacking C-terminal tyrosines. In each case, equivalent immunoprecipitation of CoSrc kinases was verified by anti-FLAG Western blotting (data not shown).

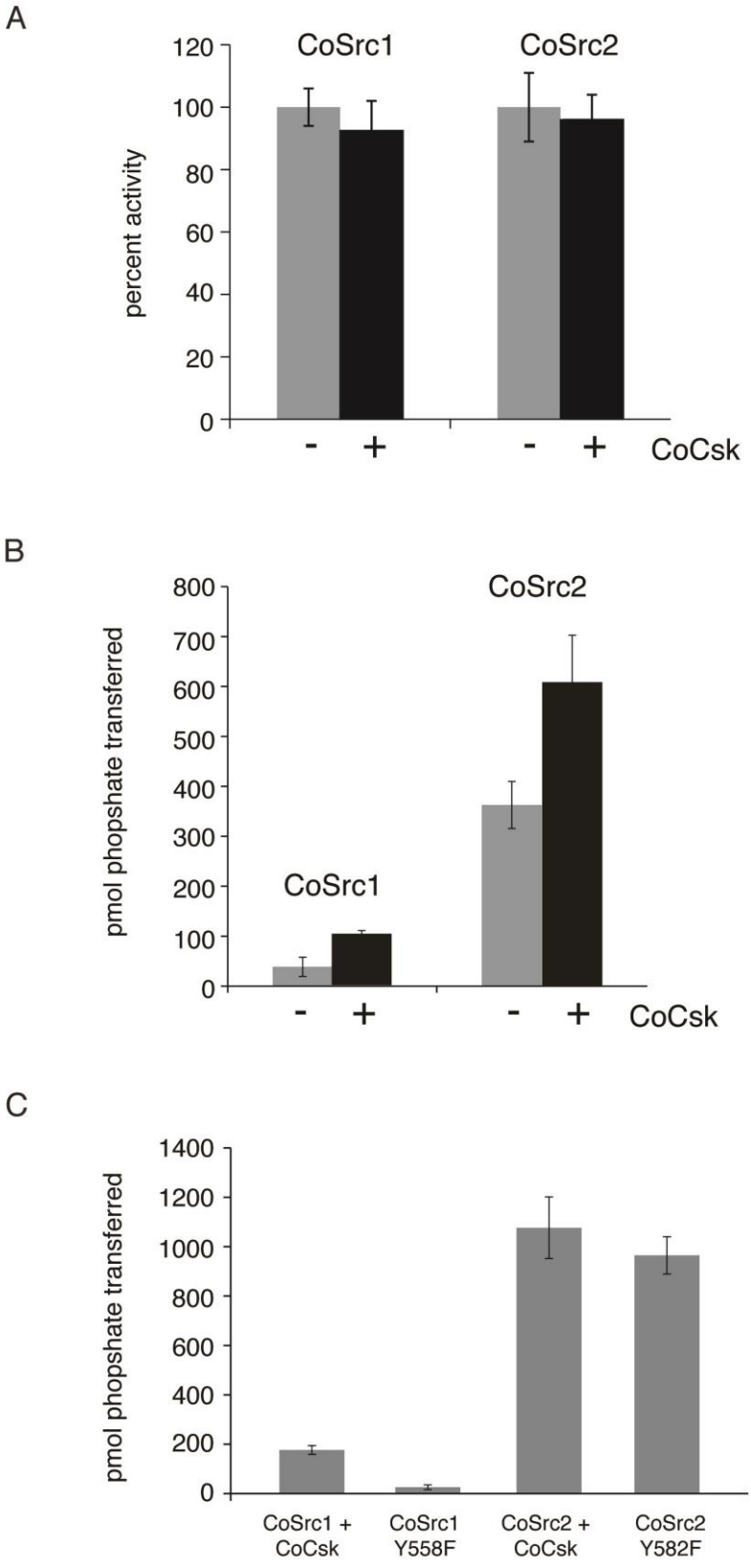

A
CoSrc1
CoSrc2
percent activity
120
100
80
60
40
20
0
- + - + CoCsk
B
CoSrc2
CoSrc1
pmol phopshate transferred
800
700
600
500
400
300
200
100
0
- + - + CoCsk
C
pmol phopshate transferred
1400
1200
1000
800
600
400
200
0
CoSrc1 + CoCsk
CoSrc1 Y558F
CoSrc2 + CoCsk
CoSrc2 Y582F

Figure 3-14. **CoCsk does not inhibit CoSrc activity in SYF cells.** Western blot analyses of CoSrc1 (left) and CoSrc2 (right) activity (wild-type or C-terminal tail YF mutants) in the presence or absence of CoCsk. Whole cell lysates of SYF cells were analyzed by anti-pTyr blotting. The filled arrow denotes the position of CoSrc1; the empty arrow denotes the position of CoSrc2. Equal expression of CoSrc1 and CoSrc2 was confirmed by anti-FLAG blotting and equal expression of CoCsk by anti-V5 blotting. Quantitation was performed as described in the legend of Figure 3-10.

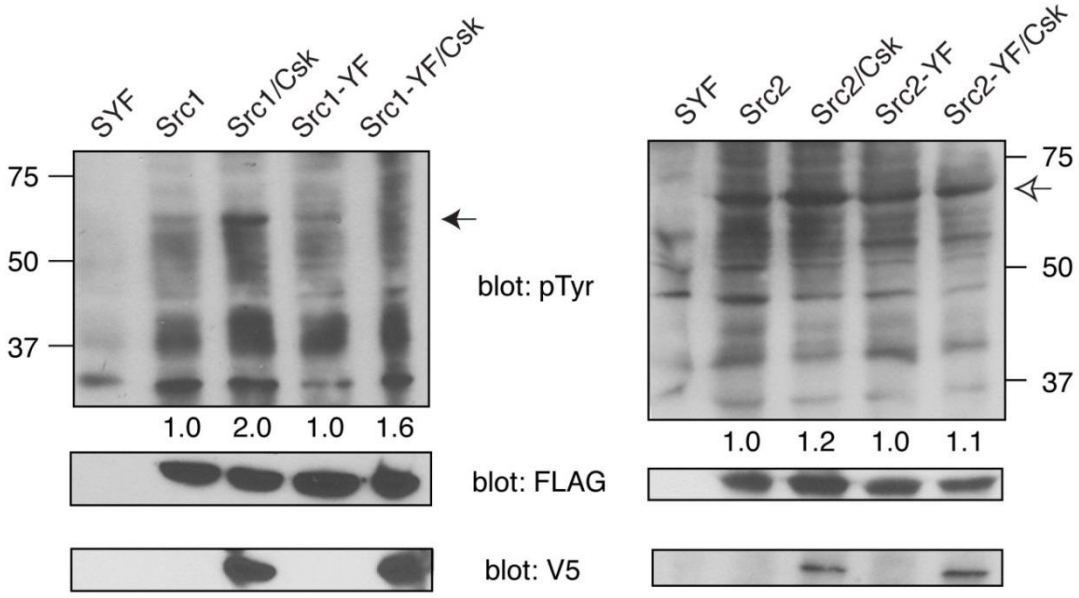

SYF
Src1
Src1/Csk
Src1-YF
Src1-YF/Csk
75
50
37
1.0
2.0
1.0
1.6
blot: pTyr
blot: FLAG
blot: V5
SYF
Src2
Src2/Csk
Src2-YF
Src2-YF/Csk
75
50
37
1.0
1.2
1.0
1.1

Figure 3-15. **CoSrc localization in *C. owczarzaki* cells.** Cells were fixed and immunostained with CoSrc1-, CoSrc2-, or actin-specific antibodies prior to fluorescence microscopy. Lectin stains the entire cell as well as the filopodia.
Localization experiments performed by Dr. Hiroshi Suga, University of Barcelona.

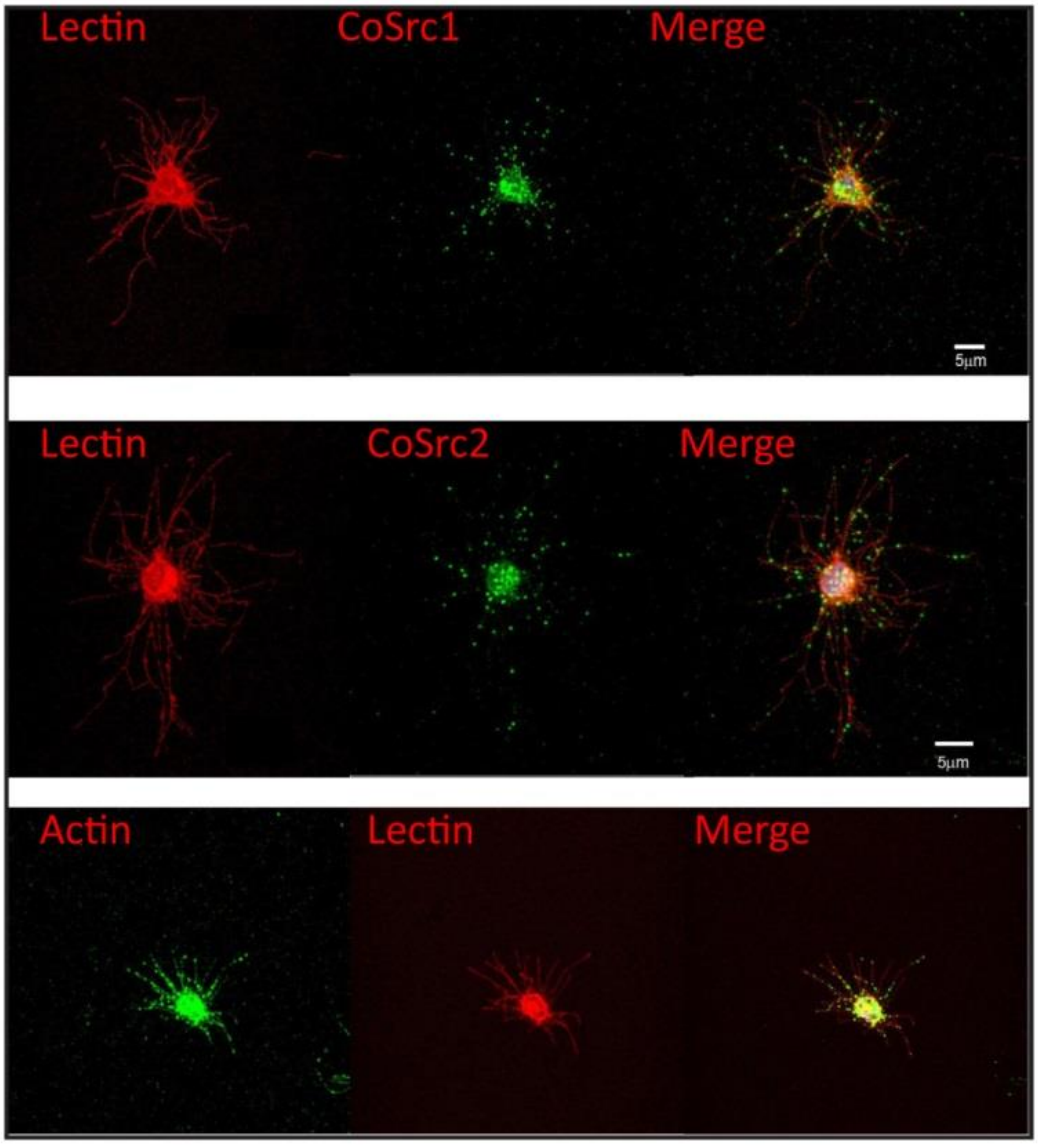
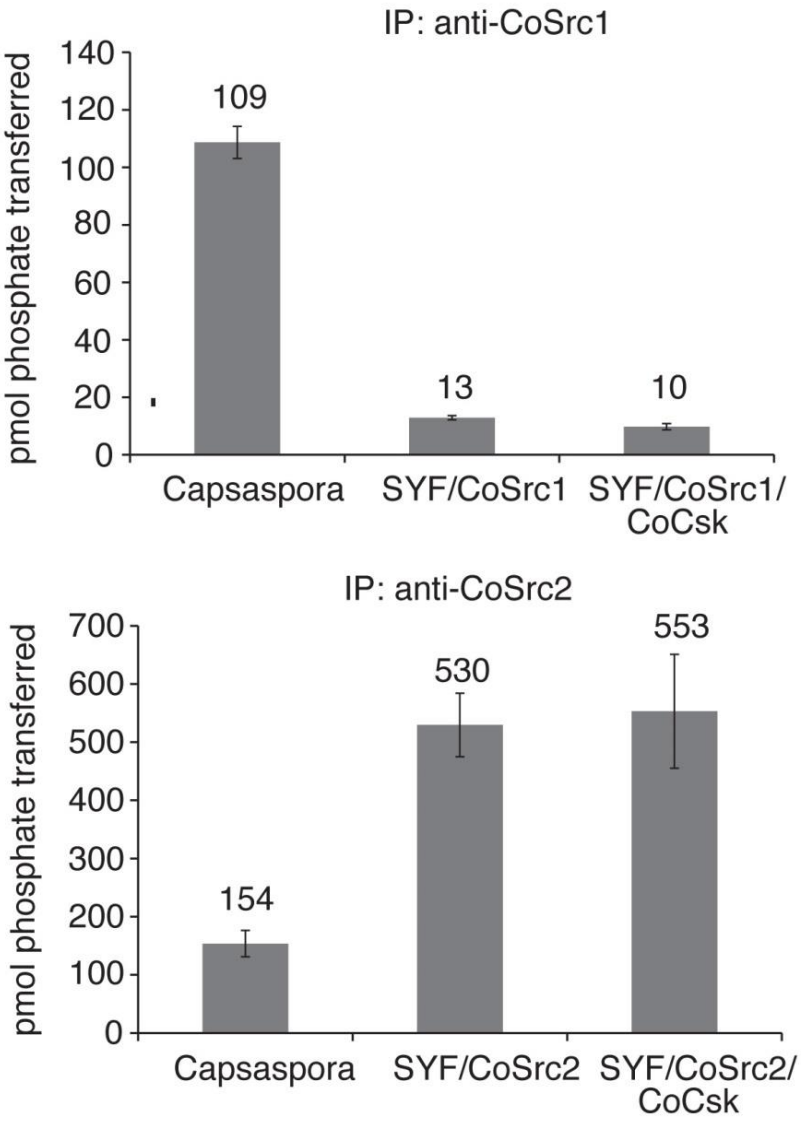
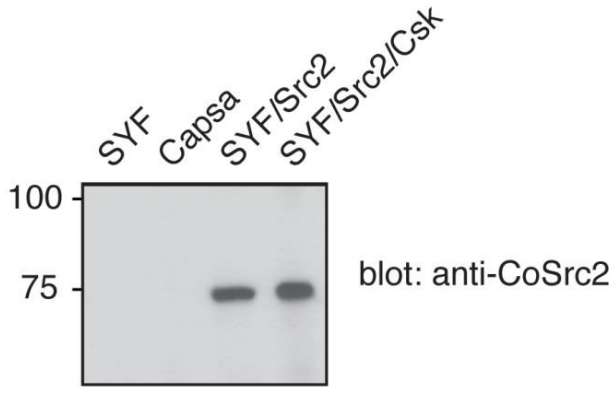

Lectin
CoSrc1
Merge
5µm
Lectin
CoSrc2
Merge
5µm
Actin
Lectin
Merge

Figure 3-16. **CoSrc activity in *C. owczarzaki* cells.** (A) Anti-CoSrc1 immunoprecipitation reactions were conducted with lysates from *Capsaspora* cells or SYF cells overexpressing CoSrc1 in the presence or absence of CoCsk. Activity measurements were performed by the phosphocellulose paper assay with the Src synthetic peptide and [γ-^{32}P]ATP . (B) Anti-CoSrc2 immunoprecipitation reactions were conducted with lysates from *Capsaspora* cells or SYF cells overexpressing CoSrc2 in the presence or absence of CoCsk. (C) Representative Western blot to compare the levels of CoSrc2 in *Capsaspora* and SYF cells.


IP: anti-CoSrc1
pmol phosphate transferred
140
120
100
80
60
40
20
0
109
13
10
Capsaspora
SYF/CoSrc1
SYF/CoSrc1/
CoCsk
IP: anti-CoSrc2
pmol phosphate transferred
700
600
500
400
300
200
100
0
154
530
553
Capsaspora
SYF/CoSrc2
SYF/CoSrc2/
CoCsk


SYF
Capsa
SYF/Src2
SYF/Src2/Csk
100
75
blot: anti-CoSrc2

Chapter 4

Lack of Csk-mediated regulation of Src in *Ministeria vibrans*

Abstract

The small marine protozoan *Ministeria vibrans*, like *Capsaspora owczarzaki*, is a member of the filasterean group, a sister clade to the choanoflagellates. As in *C. owczarzaki*, *M. vibrans* contains two Src homologues and one Csk (MvSrc1, MvSrc2 and MvCsk). MvSrc1 and MvSrc2 possess the conserved domain arrangement seen in mammalian Src family kinases. Our experiments on MvSrc1 and MvSrc2 showed them to be active kinases, with properties similar to CoSrc1 and CoSrc2. Interestingly, MvSrc2 was able to phosphorylate a serine kinase substrate as well as a tyrosine kinase substrate, suggesting that it may have served as an intermediary kinase between the emergence of tyrosine-specific kinases and the more ancient serine/threonine kinases. Just as CoCsk had no enzymatic activity and did not down-regulate CoSrc, we found that MvCsk is also a dead kinase. When a critical residue in the kinase domain of both CoCsk and MvCsk is mutated to its conserved mammalian amino acid, both kinases gain partial recovery of activity. Presumably this, and most likely other amino acid changes, evolved later in evolutionary history, allowing the complex regulatory relationship between Src and Csk to emerge. The lack of proper Src regulation appears to be a hallmark of the filasterean lineage.

Introduction

The transition from unicellular organisms to complex, multicellular animals has remained a poorly understood process [76, 77]. The first identification of tyrosine kinases outside of the metazoan clade was made in choanoflagellates, the closest known relatives to multicellular animals [80, 81, 83]. We are particularly interested in studying unicellular Src and Csk. In mammalian cells, Csk phosphorylates Src on its C-terminal tail, stabilizing the autoinhibited, closed form of the kinase. Src is found to be over-expressed and highly activated in a wide variety of human cancers, including colon, breast, pancreas and brain, highlighting the importance of proper down-regulation [13]. MbSrc1 from the choanoflagellate *M. brevicollis* had similar activity to its mammalian counterpart. However, MbCsk, although phosphorylating MbSrc on its C-terminal tail, did not down-regulate Src activity [93].

To further investigate the role of tyrosine kinases in unicellular organisms, we studied the closest relative of both Metazoa and choanoflagellates, the filasterean *Capsaspora owczarzaki* (Chapter 3). We found that although its Src proteins were active and had similar properties to that of their mammalian homologues, Csk had not yet evolved its role of Src regulation. Unlike MbCsk, CoCsk was unable to even phosphorylate its respective Src kinases, let alone affect their activities. In order to determine if the entire filasterean lineage shared this lack of Csk-mediated regulation of Src, we turned to *Ministeria vibrans*, a small single-celled eukaryote that completes the Filasteria clade along with *Capsaspora*. *M. vibrans* is a bacteria-eating marine protist, which is characterized by radiating tentacles around a central body Figure (1-6) [97]. A cDNA library from both aggregated and dispersed *Ministeria* cells was created to represent the most diverse gene collection.

A PCR-based survey was conducted in *Ministeria vibrans*, identifying 15 tyrosine encoding genes. Of these, 7 are classified as receptor tyrosine kinases, while the other 8 are nonreceptor tyrosine kinases. The nonreceptor tyrosine kinases were further subdivided into 6 families, most of them representing groups common in both *C. owczarzaki* and mammalian cells (Src, Csk, Abl, Fak, Fes, and one CTK1 family that consists of a tyrosine kinase region

without any other modular domains). The only receptor tyrosine kinase family in *M. vibrans* is a unique enzyme designated RTK1, which does not share domain architecture with any other known RTK family, including those of the closely-related *C. owczarzaki* [98].

Here we report that the *Ministeria* genome contains two Src-like kinases (MvSrc1 and MvSrc2) as well as one Csk homologue (MvCsk). We have cloned, expressed and characterized the proteins in order to compare them to their mammalian counterparts, and those in *C. owczarzaki*. MvSrc1 and MvSrc2 have a similar domain arrangement as mammalian and *C. owczarzaki* Src, with the conserved regulatory tyrosine residues. Their substrate specificity is similar to mammalian c-Src, though MvSrc2 displays activity towards a serine/threonine peptide substrate as well. This serine/threonine activity was not seen in *C. owczarzaki* CoSrc1 or CoSrc2, and may be indicative of a more primitive kinase than has yet been studied, possessing dual-specificity to phosphorylate both tyrosine and serine/threonine. MvSrc1 and MvSrc2 are able to autophosphorylate, presumably on their activation loop tyrosine, and this autophosphorylation is linked to enzymatic activity. Like CoSrc1 and CoSrc2, MvSrc displays SH3-dependent (but not SH2-dependent) substrate targeting.

As seen in *Capsaspora* CoCsk, MvCsk displays no activity toward a general tyrosine kinase substrate or purified MvSrc1 or MvSrc2. From our studies of *C. owczarzaki* and *M. brevicollis*, it appears that Csk-mediated regulation of Src arose later at the onset of metazoan evolution. However, the complete lack of Csk enzymatic activity seems to be a common trend in the Filasterean clade. The tight regulation of Src was most likely imperative in the evolution of cell-cell communication in multicellular organisms.

Results

Cloning MvSrc1 and MvSrc2 from *M. vibrans*. The genome of *M. vibrans* contains two Src-like kinases, which we have designated MvSrc1 and MvSrc2. Both *M. vibrans* Src kinases have the same basic domain structure as *Monosiga brevicollis* MbSrc1, *Capsaspora owczarzaki* CoSrc1 and CoSrc2, and mammalian Src: SH3, SH2, and kinase catalytic domains (Figure 4-1). The two tyrosine residues important for regulation are conserved in the *M. vibrans* Src kinases. MvSrc1 and MvSrc2 contain a tyrosine (Tyr-435 and Tyr-478, respectively) in the activation loop, corresponding to the c-Src autophosphorylation site. In mammals, phosphorylation of this site results in increased activity. The tyrosine in the C-terminal tail is also conserved in MvSrc1 and MvSrc2 (Tyr-546 and Tyr-589, respectively). When Csk phosphorylates this site in c-Src, the autoinhibited form of the enzyme is promoted. Although Csk is active in the choanoflagellates *Monosiga ovata* and *Monosiga brevicollis*, it does not induce autoinhibition in their corresponding Src. On the other hand, Csk has no enzymatic activity in *Capsaspora owczarzaki*.

MvSrc1 and MvSrc2 Function as Active Kinases- We expressed and purified MvSrc1 and MvSrc2 using the Sf9/baculovirus system. We determined kinase activity toward a synthetic Src substrate peptide. Both enzymes were active in a concentration-dependent manner, with MvSrc1 displaying approximately 10-fold higher activity at all protein concentrations (Figure 4-2). We then investigated substrate specificity using synthetic peptides containing substrate motifs for a variety of tyrosine kinases. MvSrc1 showed results resembling that of both mammalian and *Capsaspora owczarzaki* Src kinases, with the highest activity toward the Src peptide substrate (Figure 4-3). While MvSrc2 had the greatest preference for the Src substrate as well, it also displayed a significant amount of activity toward the protein kinase A substrate peptide (Kemptide, LRRASLG), a serine/threonine kinase substrate (Figure 4-3). The activity of MvSrc2 toward the protein kinase A substrate showed a dependence on peptide concentration (Figure 4-4). We were unable to achieve saturation in these studies, but the K_m value is in the millimolar range. Phosphorylation of the peptide was confirmed by MALDI-TOF (data not shown). The ability of a Src kinase to phosphorylate a serine/threonine substrate has not been observed before in mammalians, or any of the unicellular eukaryotes previously studied.

As well as playing an important role in Src family kinase autoinhibition, the SH2 and SH3 domains function in substrate recognition. This has been shown in *in vitro* assays; mammalian Src will phosphorylate a synthetic peptide containing an SH2 or SH3 binding domain at higher levels [29, 106]. MvSrc1 and MvSrc2 both have enhanced activity toward a peptide with an SH3 ligand, but show no discernible preference for a peptide that contains an SH2 ligand (Figure 4-5). MbSrc1 and MbSrc2 from *M. brevicollis* and CoSrc1 and CoSrc2 from *C. owczarzaki* show the same pattern of substrate targeting as *M. vibrans* Src1 and 2.

Autophosphorylation on the activation loop is a hallmark of Src family kinases. In mammalian Src, this modification increases kinase activity by destabilizing the autoinhibited conformation. To test whether MvSrc1 and MvSrc2 can autophosphorylate, we dephosphorylated the purified proteins by treating them with YOP tyrosine phosphatase. We then incubated them with [γ-^{32}P]ATP, stopping reactions at time points between 0 and 10 minutes. The reactions were separated by SDS-PAGE and analyzed by autoradiography. Both MvSrc1 and MvSrc2 displayed the ability to autophosphorylate (Figure 4-6A). To we also tested the effect of autophosphorylation on enzymatic activity. We dephosphorylated MvSrc1 and MvSrc2 and then allowed autophosphorylation to occur as above. Activity was observed toward a synthetic Src peptide using the phosphocellulose paper binding assay. Dephosphorylated MvSrc1 and MvSrc2 had no detectable activity, while the ATP-treated enzymes were active (Figure 4-6B). To investigate the activities of MvSrc1 and MvSrc2 in a cellular environment, we expressed them in mammalian Src/Yes/Fyn-deficient fibroblasts. The cells lack Src family kinases, offering cellular context with low background phosphotyrosine levels [107]. FLAG-tagged MvSrc1, MvSrc2, or c-Src were expressed, and overall tyrosine kinase activity was measured by Western blotting with an anti-phosphotyrosine antibody. Mammalian c-Src showed robust autophosphorylation and phosphorylation of additional SYF cell proteins (as expected), but little activity could be seen for MvSrc1 or MvSrc2 (Figure 4-7).

Cloning *M. brevicollis* MvCsk. A C-terminal Src kinase (Csk) homologue was identified from the *M. vibrans* genome. Its sequence was compared to the Csk homologues from humans, *Xenopus* (frog), *E. fluviatis* (a simple sponge), *M. brevicollis* and *C. owczarzaki* (Figure 4-8). MvCsk possesses the conserved SH2, SH3, and kinase domain arrangement. MvCsk contains

the key catalytic residues discussed in Chapter 3, including all six residues previously identified as being important in mammalian Src recognition (Figure 3-7).

MvCsk Has No Measurable Activity. Full-length MvCsk was cloned and expressed in bacterial cells as a fusion protein with glutathione S-transferase (GST). GST-tagged MvCsk protein was purified by glutathione-agarose chromatography. MvCsk activity was assayed with a general tyrosine kinase substrate, the synthetic peptide poly(Glu_4EY). Similarly to the *Capsaspora* Csk, MvCsk had no activity toward its substrate (Figure 4-9). Next, we tested the activity of MvCsk toward purified MvSrc1 or MvSrc2. The activity of mammalian Csk toward a kinase dead truncated form of c-Src was also tested as a positive control for the assay. Purified proteins were incubated with [γ-^{32}P]ATP and separated by SDS-PAGE. No MvCsk activity toward the *Ministeria* Src proteins could be detected by autoradiography, although mammalian Csk phosphorylates c-Src under these conditions (Figure 4-10).

Both CoCsk and MvCsk show strong sequence homology to their counterparts in mammalians and choanoflagellates. The important residues for ATP binding and catalysis are conserved in the kinase domain. However, the structure of Csk is very different from Src family kinases. While an isolated Src kinase domain is extremely active, the Csk kinase domain requires interactions with its SH2 and SH3 regulatory domains in order to function [59]. Residues in the SH3-SH2 linker, as well as a β3-αC loop in the kinase domain that must interact with a portion of the SH2 domain, have been shown necessary for the activation of mammalian Csk [109, 110]. In Chapter 4 we discussed an amino acid mutation in the SH2-kinase interface of *Capsaspora* Csk that would partially restore activity kinase activity when mutated. In mammalian, *Xenopus*, sponge and *M. brevicollis* Csk, this residue is an asparagine or aspartic acid residue, but in the inactive *Capsaspora* and *Ministeria* Csk, there is a glycine present. We produced a G226N mutant form of MvCsk, expressed and purified the GST fusion protein, and tested activity against a synthetic substrate peptide. The mutation caused a partial recovery of MvCsk activity toward the Src substrate peptide (Figure 4-11). The G226N mutation was not able to restore MvCsk activity towards purified MvSrc1/2 as seen through autoradiography (data not shown). These results recapitulate the experiments from CoCsk, and reinforce the importance of the SH2-kinase domain interactions in Csk activity.

MvCsk Does Not Negatively Regulate MvSrc. Mammalian Csk phosphorylates the C-terminal tail of Src, stabilizing the autoinhibited form and reducing its kinase activity. MvCsk did not display any measurable activity towards a general tyrosine kinase substrate (Figure 4-9) or MvSrc (Figure 4-10). Next, we tested its ability to regulate MvSrc activity. MvSrc1 and MvSrc2 were incubated with MvCsk *in vitro*, and MvSrc activity was measured in a radioactive kinase assay. No inhibition of MvSrc1 or MvSrc2 was observed in the presence of MvCsk (Figure 4-12). Instead, MvSrc activity was modestly increased by MvCsk treatment. A similar result was obtained for the *Capsaspora owczarzaki* Src and Csk enzymes when CoSrc1 and CoSrc2 were co-expressed with CoCsk in mammalian cells (Figure 3-13B).

Discussion

Studying the unicellular counterparts to mammalian signaling components can help us to better understand the evolutionary process. The closest known relatives to multicellular organisms, Choanoflagellates, contain both Src and Csk proteins, which have similar enzymatic activity as in metazoans. Although Csk from *M. brevicollis* is able to phosphorylate Src, it does not impose negative regulation [93]. In *Capsaspora owczarzaki*, the next closest relative to both metazoans and Choanoflagellates, Src is enzymatically functional, yet Csk lacks any catalytic activity. To see whether this trend also exists in *Ministeria vibrans*, another organism in the same phylogenetic group as *Capsaspora*, we investigated the activity of the two Src-like kinases and one Csk homologue present in its genome.

The activities of *Ministeria* Src1 and Src2 are comparable to mammalian c-Src (Figure 4-2). MvSrc1 had an approximately 10-fold higher specific activity than MvSrc2 toward a synthetic peptide containing an optimal Src recognition sequence (Figure 4-3). In addition, both phosphorylated an SH3 ligand-containing peptide at a higher rate as compared to a control peptide, but did not show a preference for a peptide with an SH2 ligand (Figure 4-5). These

results mirror those obtained for the *C. owczarzaki* CoSrc1 and 2, as well as *M. brevicollis* MbSrc1, strengthening the idea that the coupling of the SH2 and kinase domains and autoregulation developed late in SFK evolution.

Eukaryotic signaling is dominated by serine, threonine, and tyrosine phosphorylation by the eukaryotic protein kinase (ePK) superfamily [83]. Protein tyrosine phosphorylation is thought to be a more recent development in evolutionary history, previously thought to be metazoan-specific. Protein tyrosine kinases (PTKs) normally specifically phosphorylate tyrosine residues, while other ePKs phosphorylate mainly serine and threonine residues [111]. Jensen has proposed that primitive enzymes in general had broader specificities than modern enzymes, allowing fewer enzymes to carry out the processes needed for ancestral organisms [112]. Gene duplication and divergence led to more specialized genes, leading to the creation of enzyme families and superfamilies. The development of tyrosine-specific kinases from serine/threonine kinases is thought to have occurred through a similar process. A few protein kinases, such as Wee1, VRK3, and STRADα, designated dual-specificity kinases, have the ability to phosphorylate both serine and tyrosine residues [113, 114]. Dual-specificity kinases most likely served as an intermediary step between the more ancient serine/threonine kinases and the more recent tyrosine kinases. It is hypothesized that in some cases the genes of the dual-specificity kinases were duplicated, and the new enzymes lost the ability to phosphorylate serine and threonine, leading to the birth of the new tyrosine kinase subgroup [115]. The tyrosine kinases are distinguished from these dual-specificity kinases by sequence homology and a conserved catalytic loop motif [98].

We have found that MvSrc1 and MvSrc2 both display a strong enzymatic preference for a synthetic peptide containing an optimal substrate for Src, as seen in *Capsaspora* kinases. In addition, MvSrc2 is also able to phosphorylate the protein kinase A substrate peptide (Kemptide). MvSrc1 had no activity towards the same serine/threonine kinase substrate. Based on these results, MvSrc2 may represent an intermediary form of a tyrosine kinase; it is a tyrosine kinase by amino acid sequence homology, yet it is able to act as a dual-specificity kinase, phosphorylating tyrosine as well as serine. MvSrc2 may give insight into the development of tyrosine-specific kinases from the more ancient serine/threonine kinases. More studies will need

to be done on the Src-like kinases of earlier ancestors in the metazoan lineage to test for the presence of dual specificity, or determine if the phenomenon is specific to MvSrc2.

Prior functional studies of the *C. owczarzaki* Csk revealed a lack of enzymatic activity, in contrast to all of the previously studied Csk proteins. When similar experiments were done with *Ministeria vibrans, C. owczarzaki's* closest known relative, we found that this was a similarity within the clade; MvCsk was also functionally dead. In mammalian cells, Csk is absolutely necessary to properly regulate Src activity through phosphorylation of its C-terminal tail tyrosine. As seen for CoCsk, MvCsk is unable to phosphorylate either MvSrc pure protein, or a synthetic peptide. MvCsk was also unable to show an inhibitory effect on MvSrc1/2 activity. When CoCsk was mutated at a critical residue in its kinase domain, it was able to regain partial activity towards the synthetic peptide. The corresponding mutation in MvCsk, G226N, also restored low levels of activity in a kinase assay. These results, along with the studies on CoCsk, suggest that Csk perhaps did not evolve its catalytic activity until after the split between filastereans and choanoflagellates. Alternatively, Csk may have evolved as an active enzyme that lost its activity in the filasterean lineage. If Csk initially developed as an enzymatically dead protein, it is tempting to classify the unicellular homologues as "pseudokinases." Although the Csk sequence contains all of the key residues in the kinase domain (VAIK, HRD, DFG) that are normally lacking in pseudokinases, our data is consistent with roles that have been assigned to other pseudokinases (scaffolding, activating other proteins through domain binding, or targeting of active proteins to specific cellular locations) [72, 73, 114].

Figure 4-1. **Alignment of *Monosiga brevicollis* MvSrc1 and MvSrc and *Capsaspora owczarzaki* CoSrc1 and CoSrc2.** The SH3 domain is boxed in red, the SH2 domain boxed in blue, and the kinase domain in green. Orange circles represent conserved regulatory tyrosines.

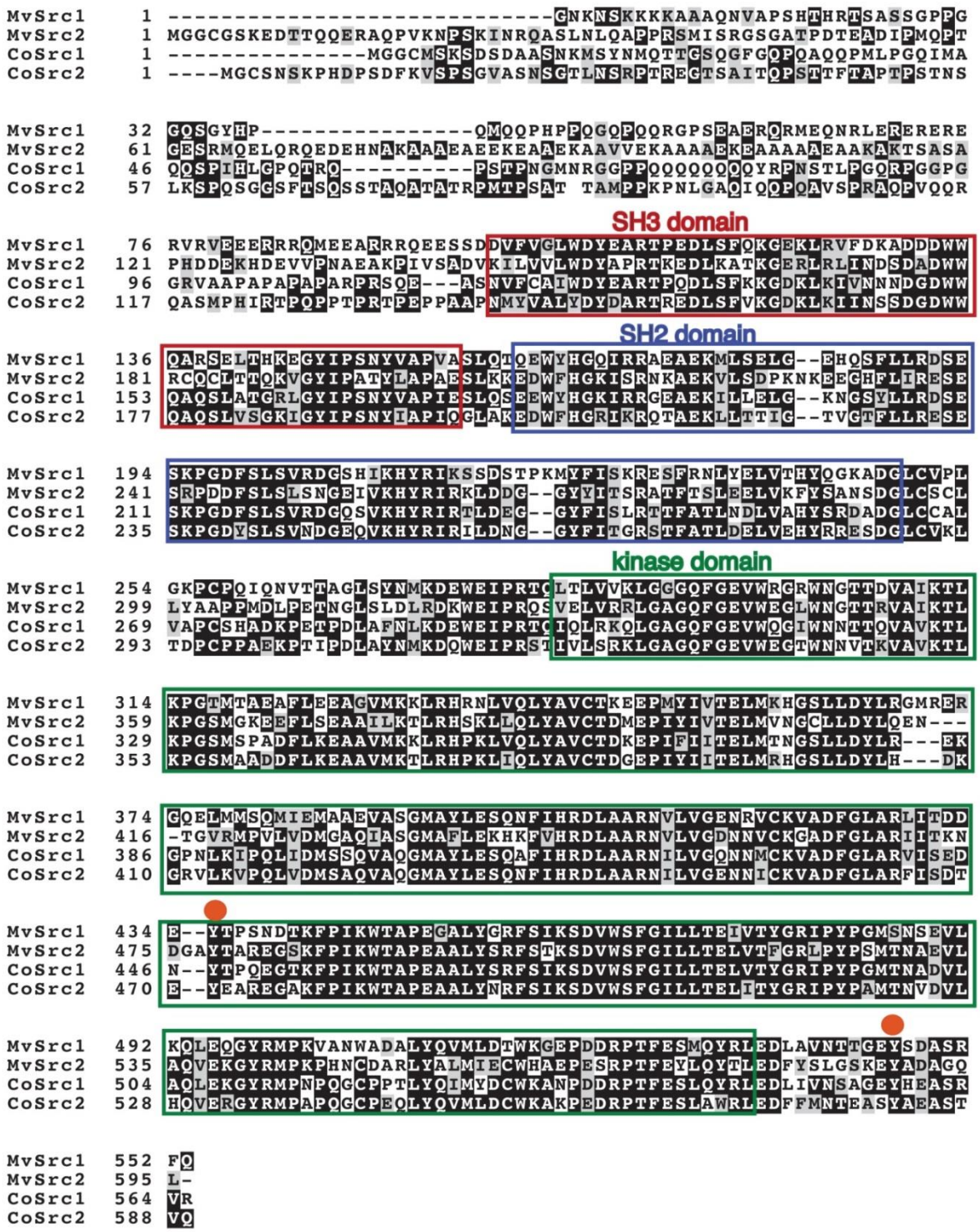
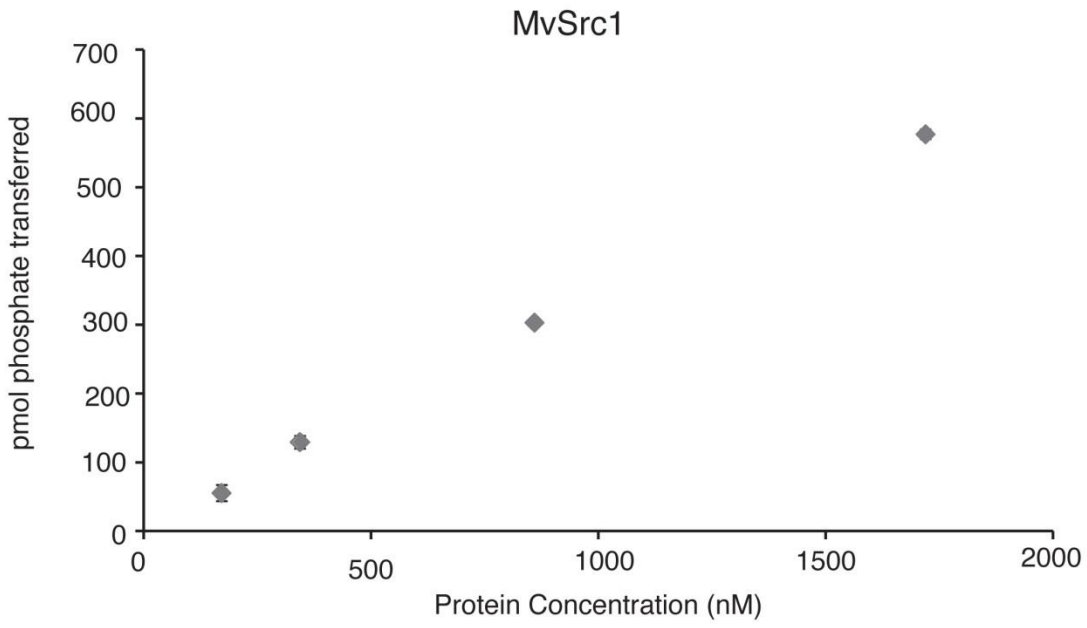

SH3 domain
SH2 domain
kinase domain

Figure 4-2. **MvSrc1 and MvSrc2 have activity toward a Src peptide.** Activity of MvSrc1 and MvSrc2 at varying enzyme concentrations. Enzyme activities were measured with 0.7 mM Src peptide substrate and 0.25 mM ATP using the phosphocellulose paper binding assay. Reactions proceeded for 9 minutes at 30°C.


MvSrc1
pmol phosphate transferred
700
600
500
400
300
200
100
0
0
500
1000
1500
2000
Protein Concentration (nM)

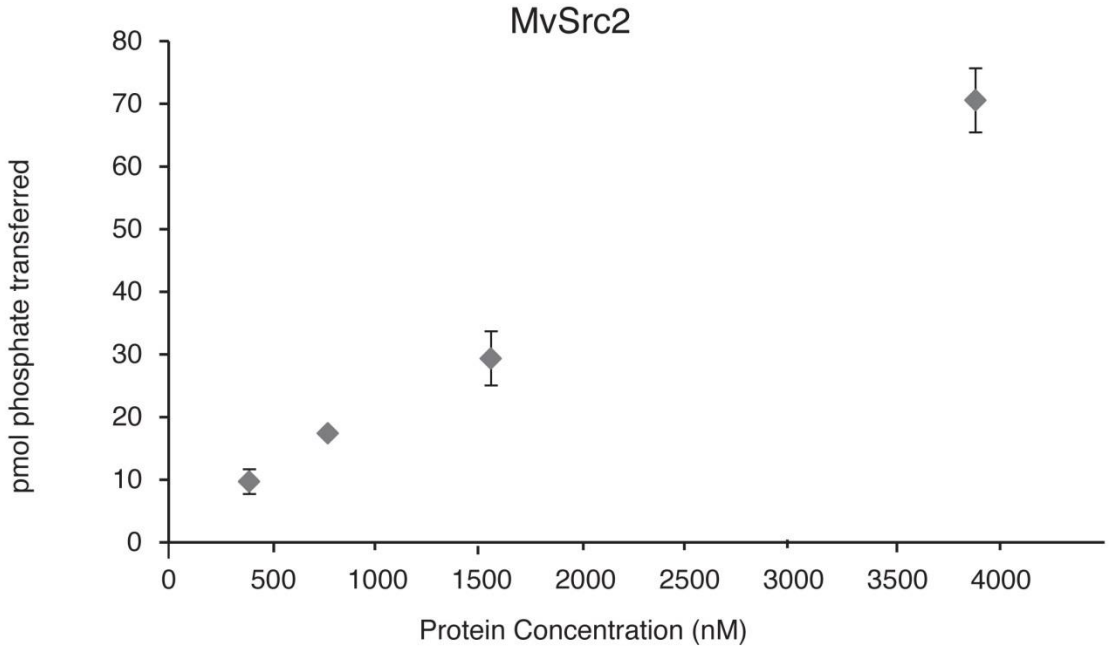
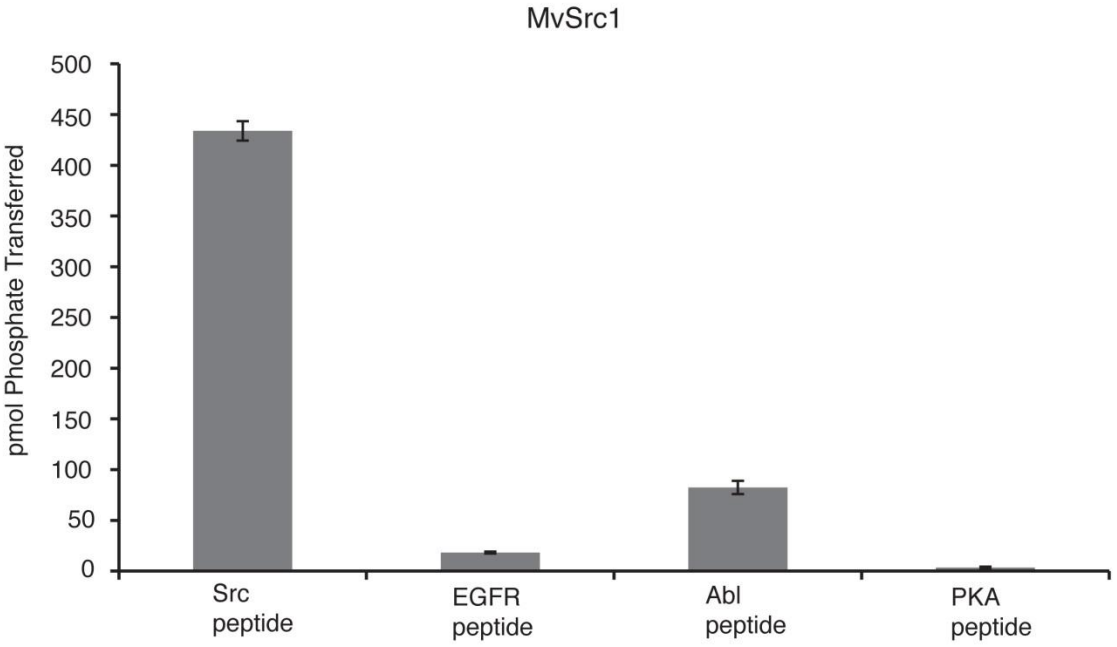

MvSrc2
pmol phosphate transferred
80
70
60
50
40
30
20
10
0
0
500
1000
1500
2000
2500
3000
3500
4000
Protein Concentration (nM)

Figure 4-3. **Substrate specificities of MvSrc1 and MvSrc2.** Substrate specificities were investigated using synthetic peptides incorporating recognition motifs from four protein kinases. Enzymes were incubated with 0.25 mM [γ-^{32}P]ATP and 800 µM peptide, and reactions proceeded for 8 minutes at 30°C.


MvSrc1
pmol Phosphate Transferred
500
450
400
350
300
250
200
150
100
50
0
Src peptide
EGFR peptide
Abl peptide
PKA peptide

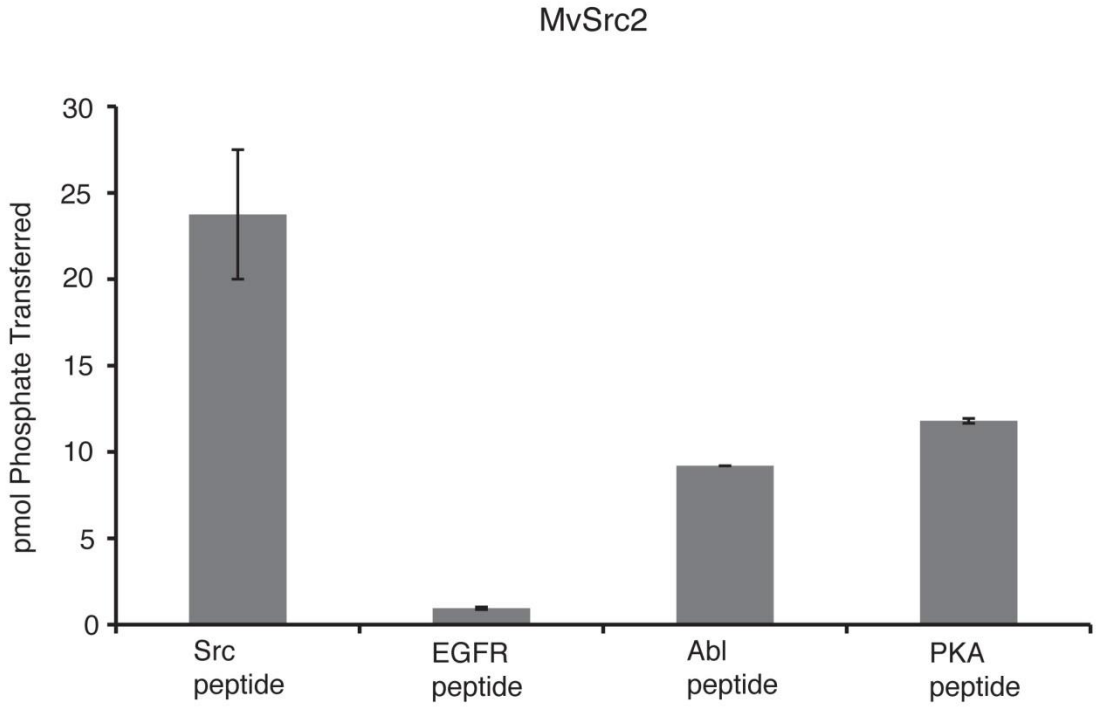

MvSrc2
pmol Phosphate Transferred
30
25
20
15
10
5
0
Src peptide
EGFR peptide
Abl peptide
PKA peptide

Figure 4-4. **MvSrc2 can phosphorylate a synthetic protein kinase A substrate.** Activity of MvSrc2 toward a PKA peptide was investigated through the phosphocellulose paper assay. Varying concentrations of peptide was used. Enzyme activity was measured with 2 µM MvSrc2 enzyme and 0.25 mM ATP. Reactions were incubated for 9 minutes at 30°C.

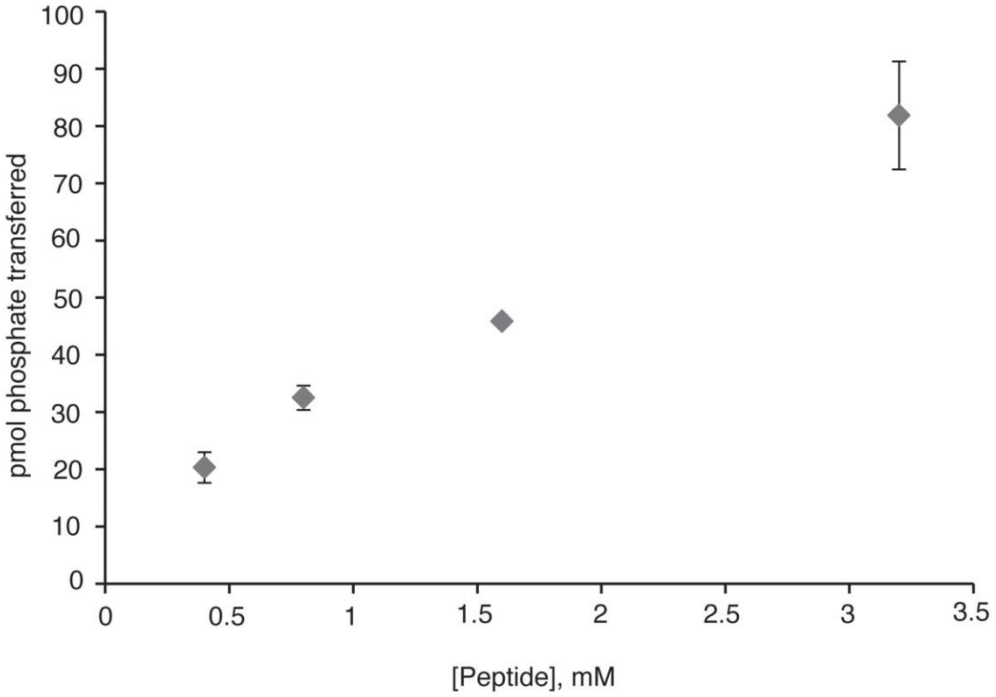

100
90
80
70
60
50
40
30
20
10
0
pmol phosphate transferred
0
0.5
1
1.5
2
2.5
3
3.5
[Peptide], mM

Figure 4-5. **Substrate targeting by MvSrc1 and MvSrc2.** MvSrc1 (1 μM) and MvSrc2 (5 μM) were tested with synthetic peptide substrates containing SH3 or SH2 ligand sequences or matched controls. Reactions proceeded for 10 min at 30°C, and mixtures were analyzed by the phosphocellulose paper assay.

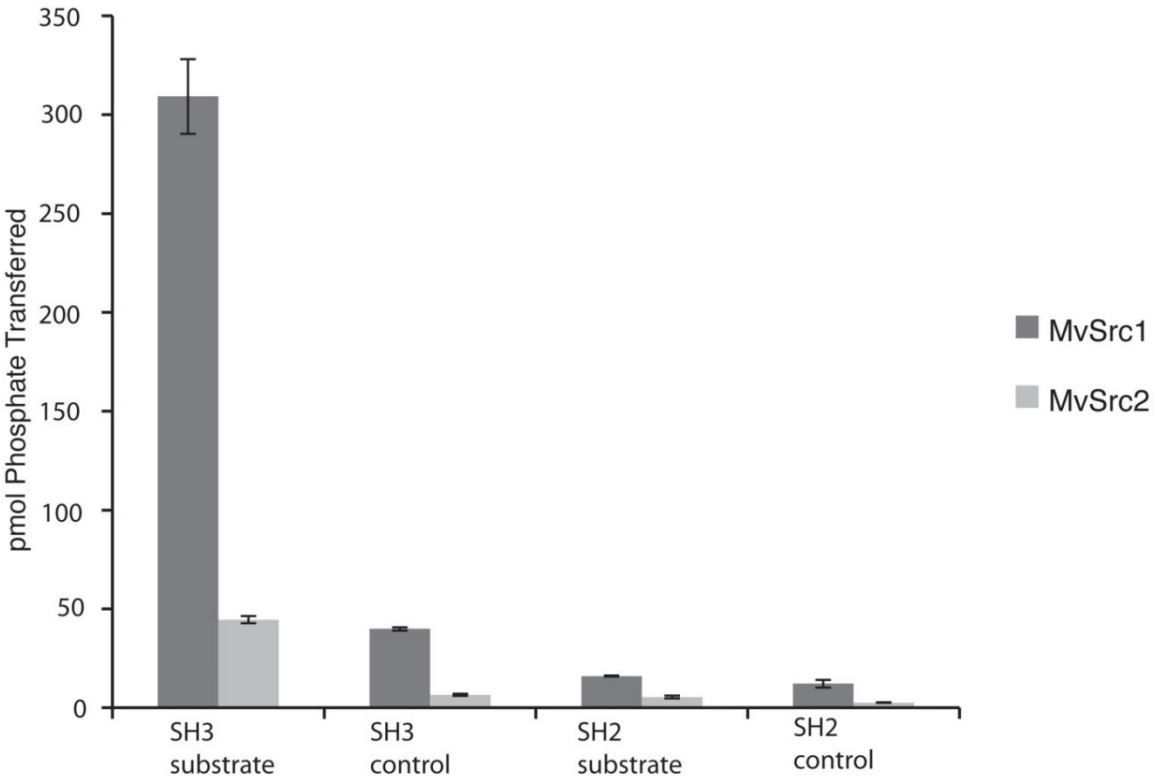
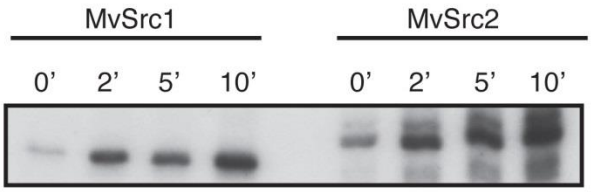
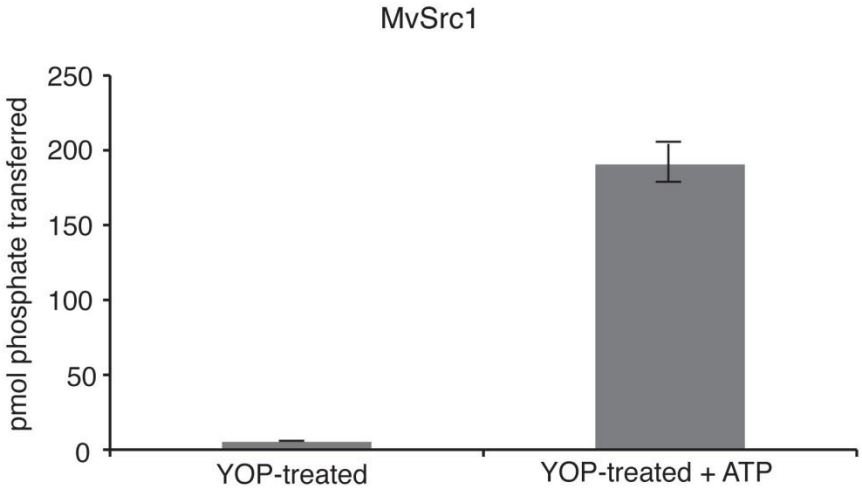

350
300
250
200
150
100
50
0
pmol Phosphate Transferred
SH3 substrate
SH3 control
SH2 substrate
SH2 control
MvSrc1
MvSrc2

Figure 4-6. **Autophosphorylation of MvSrc.** (A) MvSrc1 and MvSrc2 were treated with immobilized GST-YOP for 30 min at 30°C. The YOP was removed by centrifugation, and the mixture was incubated with [γ-^{32}P]ATP. Samples were stopped at 0, 2, 5, and 10 minute intervals with Laemmli buffer. Reactions were separated by SDS-PAGE and analyzed by autoradiography. (B) MvSrc1 and MvSrc2 were treated with immobilized GST-YOP for 30 min at 30°C. The YOP was removed by centrifugation, and the mixture was assayed directly or after incubation with ATP for 30 min at 30°C. Activity measurements were performed in duplicate with the Src synthetic peptide and the phosphocellulose paper assay

A


MvSrc1
MvSrc2
0’ 2’ 5’ 10’
0’ 2’ 5’ 10’

B


MvSrc1
pmol phosphate transferred
250
200
150
100
50
0
YOP-treated
YOP-treated + ATP

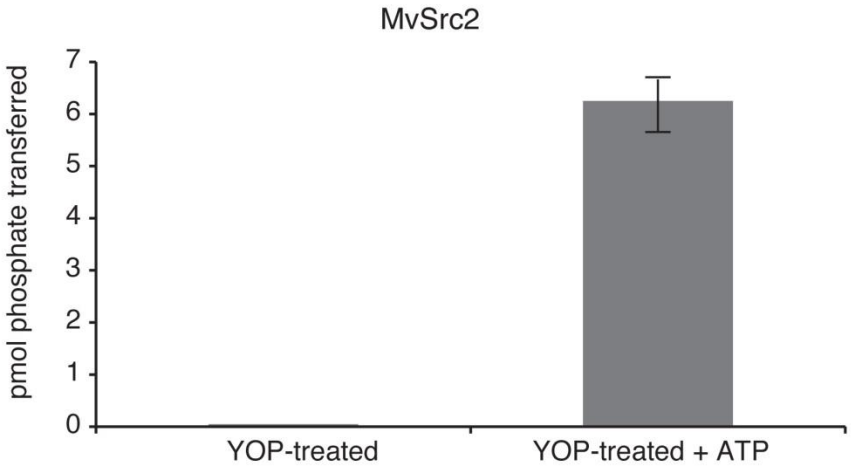

MvSrc2
pmol phosphate transferred
7
6
5
4
3
2
1
0
YOP-treated
YOP-treated + ATP

Figure 4-7. **MvSrc expression in SYF cells.** Whole cell lysates from untransfected SYF cells, or SYF cells expressing MvSrc1, MvSrc2, or mammalian c-Src, were analyzed by Western blotting with anti-phosphotyrosine antibody. The membrane was stripped and reprobed with anti-FLAG antibody to compare Src expression.

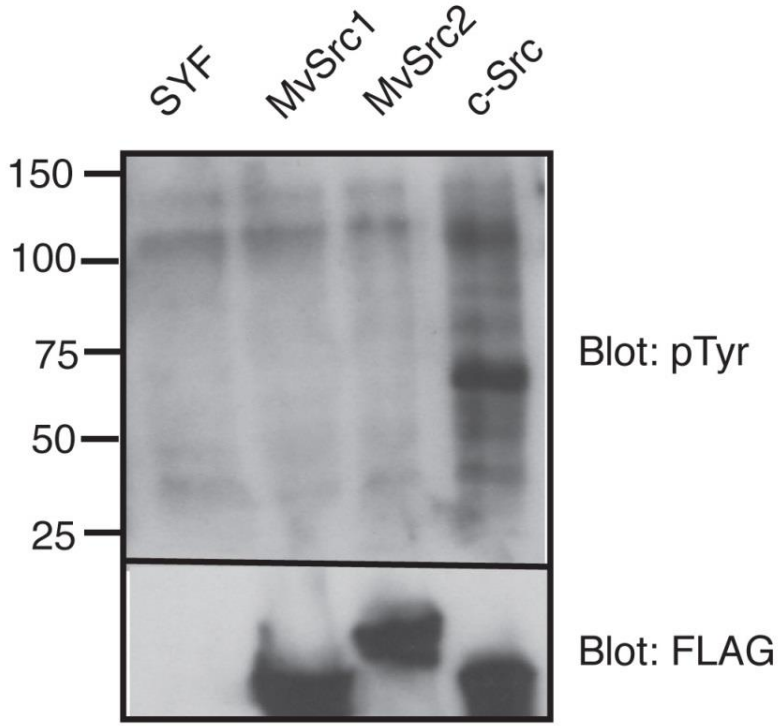

SYF
MvSrc1
MvSrc2
c-Src
150
100
75
50
25
Blot: pTyr
Blot: FLAG

Figure 4-8. **Amino acid sequence alignment of *M. vibrans* MvCsk, and human, Xenopus, Sponge, *M brevicollis,* and *C. owczarzaki* Csk kinases.** The SH3 domain is boxed in green, the SH2 is boxed in blue, and the kinase domain is boxed in purple.

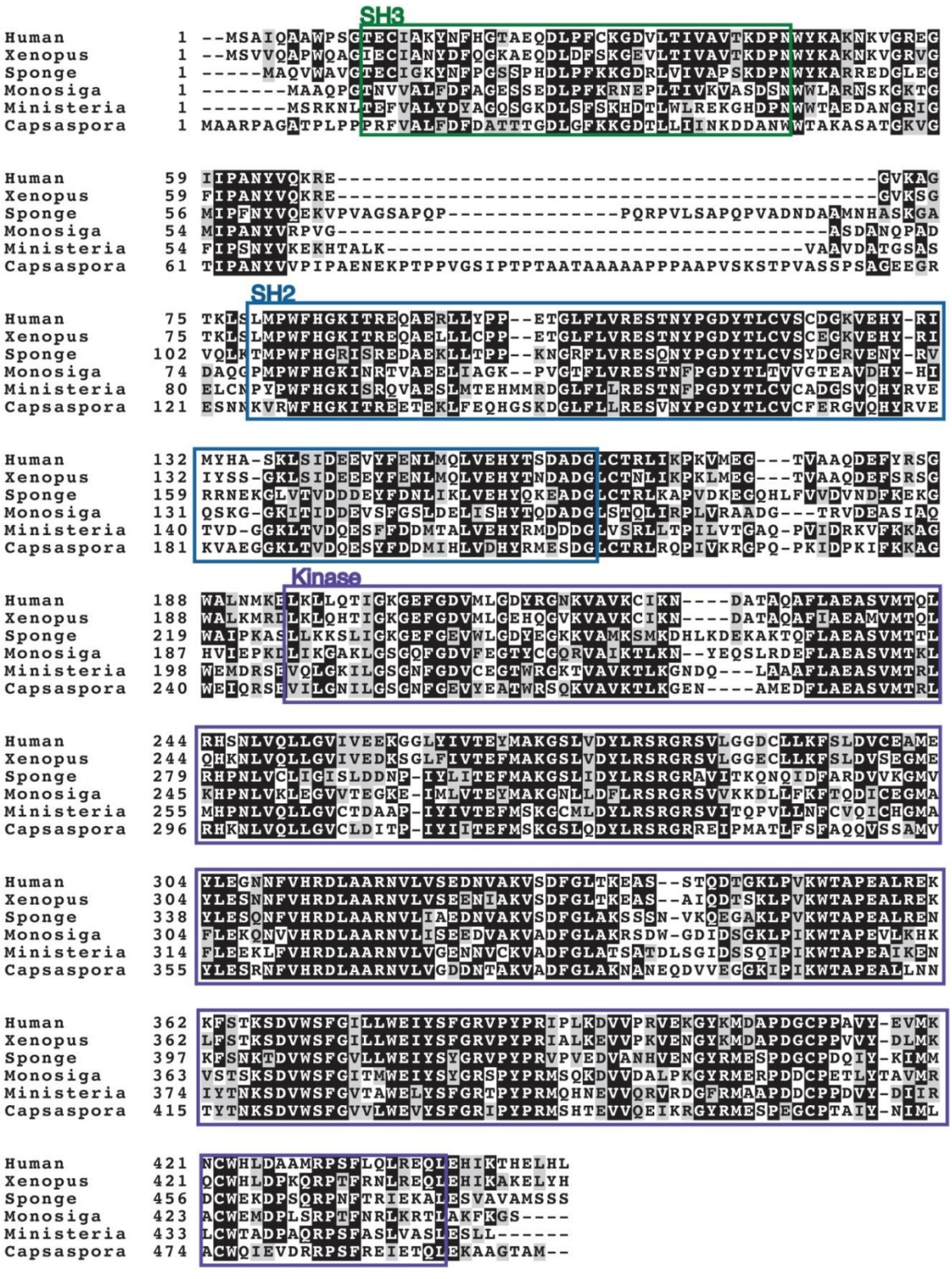

SH3
SH2
Kinase
Human
Xenopus
Sponge
Monosiga
Ministeria
Capsaspora

Figure 4-9. **Activity of MvCsk toward a synthetic peptide is undetectable.** Purified MvCsk, *M. brevicollis* MbCsk, and human Csk (mCsk) (200 nM) were assayed with poly(Glu_4Tyr) (1 mg/mL). Reactions contained 500 μM [γ-^{32}P]ATP. Activity was analyzed at various time points by scintillation counting.

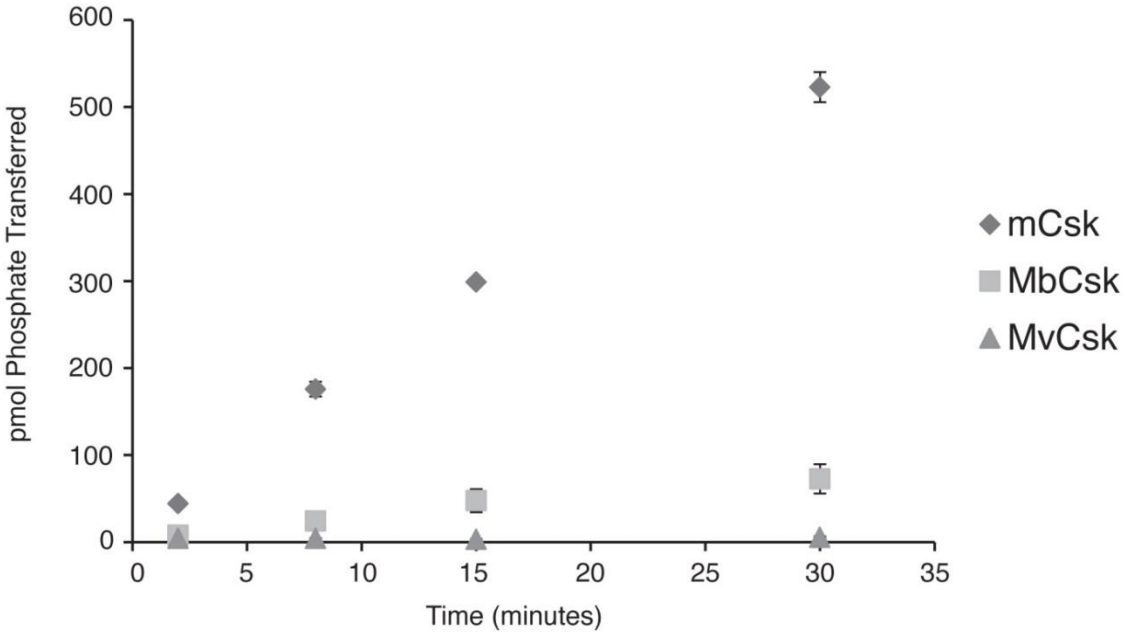
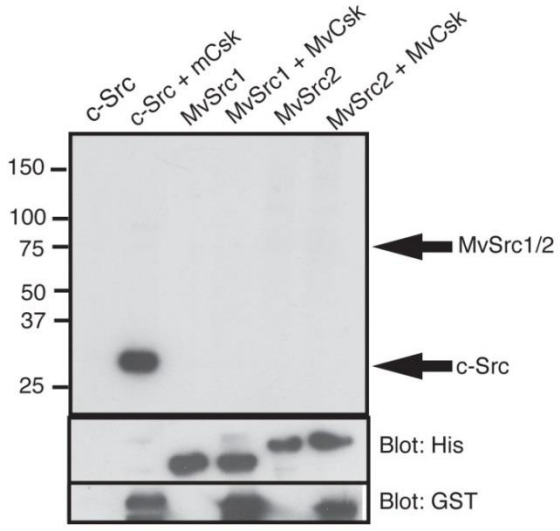

600
500
400
300
200
100
0
pmol Phosphate Transferred
0
5
10
15
20
25
30
35
Time (minutes)
mCsk
MbCsk
MvCsk

Figure 4-10. **MvCsk does not phosphorylate MvSrc in vitro.** (A) Purified MvCsk (1 μM) was incubated with purified MvSrc1 and MvSrc2 (1 μM) in the presence of [γ-^{32}P]ATP. Reactions were stopped by Laemmli buffer, and mixtures were analyzed by SDS-PAGE and autoradiography. (B) The bands from panel (A) were excised from the gel, dissolved, and analyzed by scintillation counting.

A


c-Src
c-Src + mCsk
MvSrc1
MvSrc1 + MvCsk
MvSrc2
MvSrc2 + MvCsk
150
100
75
50
37
25
MvSrc1/2
c-Src
Blot: His
Blot: GST

B

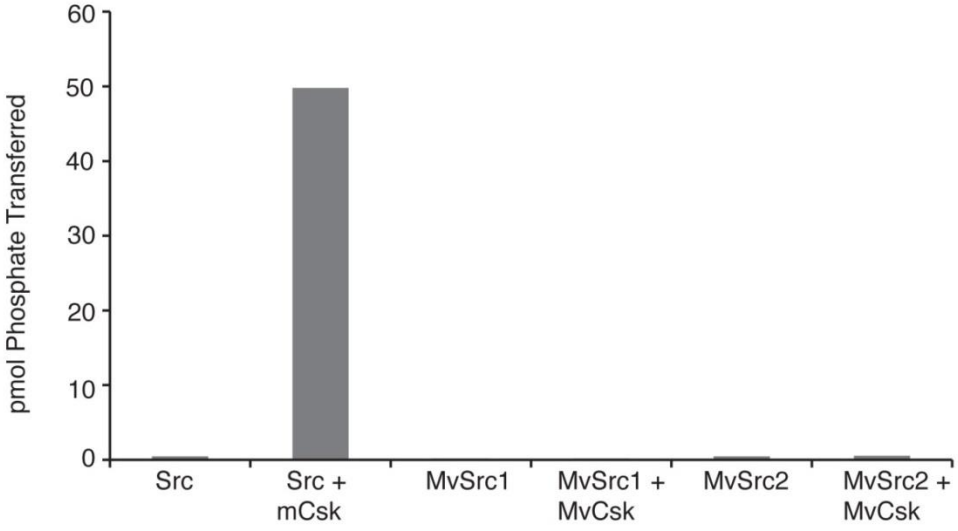

pmol Phosphate Transferred
60
50
40
30
20
10
0
Src
Src + mCsk
MvSrc1
MvSrc1 + MvCsk
MvSrc2
MvSrc2 + MvCsk

Figure 4-11. **The G226N mutation partially restores MvCsk activity.** Purified wild-type GST-CoCsk and GST-CoCsk-G226N were assayed with poly(Glu_4Tyr) (1 mg/mL). Reaction mixtures contained 500 μM [γ-^{32}P]ATP. Activity was analyzed at various time points by scintillation counting.

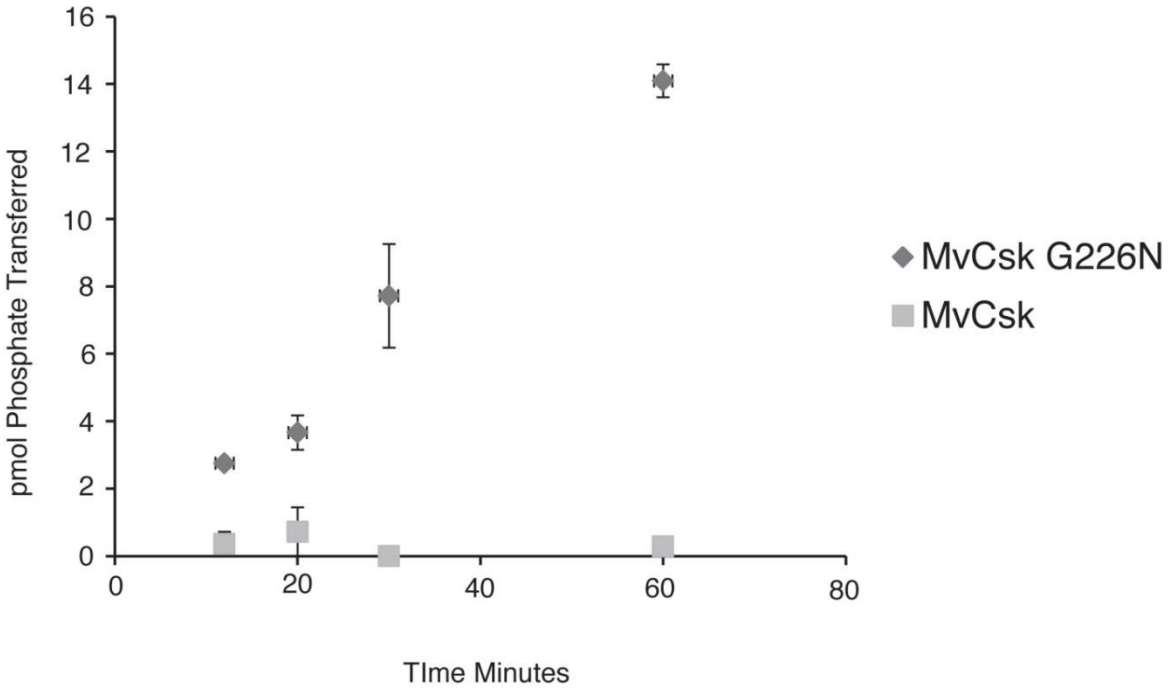
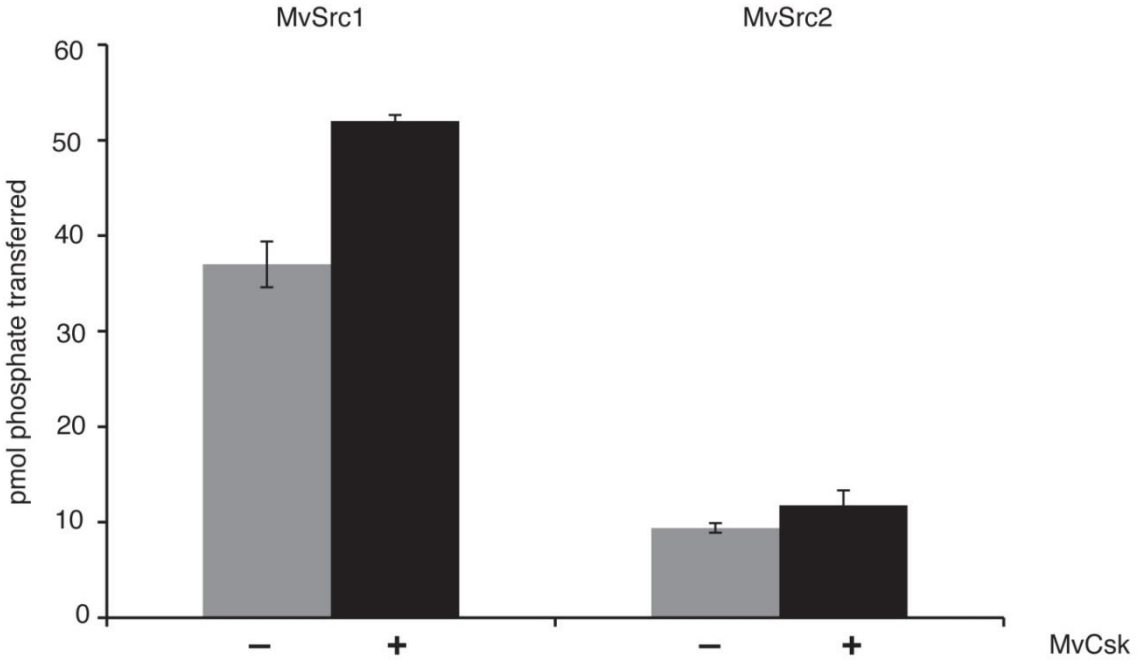

16
14
12
10
8
6
4
2
0
pmol Phosphate Transferred
0
20
40
60
80
TIme Minutes
MvCsk G226N
MvCsk

Figure 4-12. **MvCsk does not inhibit MvSrc activity.** In vitro MvSrc kinase assays were performed in the presence or absence of purified MvCsk. Reaction mixtures contained 1 μM MvSrc1 or MvSrc2, 1 μM MvCsk, 500 μM [γ-^{32}P]ATP, and 0.4 mg/mL RCM-lysozyme as a Src substrate. After 20 minutes at 30°C, reaction mixtures were analyzed by scintillation counting.


MvSrc1
MvSrc2
60
50
40
30
20
10
0
pmol phosphate transferred
–
+
–
+
MvCsk

Chapter 5

Src substrates in choanoflagellates

*Portions of this chapter were published in Schultheiss et al., BMC Biochem, **14** (2013)

Abstract

To gain more insight into the signaling mechanisms in unicellular relatives to Metazoa, we have carried out preliminary studies aimed at identifying Src substrates in choanoflagellates. The genome of *Monosiga brevicollis*, a member of the choanoflagellates, encodes 128 tyrosine kinases, of which 88 are receptor tyrosine kinases (RTKs). We have cloned, expressed and purified *Monosiga brevicollis* RTKB2. This is the first biochemical study of a unicellular receptor tyrosine kinase. RTKB2 is an enzymatically active kinase. It contains 6 repeats of a unique domain (RM2) on its cytoplasmic tail that can be phosphorylated by the Src family kinase MbSrc1. We expressed and purified an individual RM2 domain, and showed that it can be phosphorylated by MbSrc1. When phosphorylated, the RM2 domain bound to the SH2 domain of MbSrc1. In addition, we have studied a STAT homologue from the choanoflagellate *Salpingocea rosetta*. In multicellular animals, STATs act as transcription factors that are activated through phosphorylation by Jak, Src, and other tyrosine kinases. *S. rosetta* STAT contains a similar domain arrangement as mammalian STAT, sharing the conserved, functional residues. We found that when co-expressed in mammalian cells, the *S. rosetta* STAT can be phosphorylated by both c-Src and the Src-like kinase from *S. rosetta.* Peptides derived from the major phosphorylation site of *S. rosetta* STAT serve as substrates for *M. brevicollis* MbSrc1 and mammalian Src. MbSrc1 phosphorylates the *S. rosetta* STAT peptide more efficiently than its metazoan counterpart *in vitro*. Studying the signaling components of metazoan ancestors allows us to trace the development of the complex systems found in multicellular animals.

Introduction

RTKB2, a substrate for *Monosiga brevicollis* MbSrc1

Metazoan receptor tyrosine kinases (RTKs) respond to a variety of extracellular stimuli, initiating signals that regulate important cellular and developmental processes [5, 8-10]. The phosphotyrosine-based signaling system (consisting of tyrosine kinases, tyrosine phosphatases, and pTyr-binding modules) evolved relatively recently [70, 79]; pTyr signaling was originally thought to be unique to metazoans. Remarkably, recent genomic analyses have demonstrated that several unicellular organisms possess numbers of receptor and nonreceptor tyrosine kinases that are comparable to higher metazoans. Choanoflagellates are free-living aquatic protists that represent the closest unicellular relatives to metazoans [116, 117]. Tyrosine kinases are abundant in the choanoflagellates *Monosiga brevicollis*, *Monosiga ovata*, and *Salpingoeca rosetta* [81, 82, 87, 105]. Although the physiological functions of these unicellular tyrosine kinases are not yet known, they are presumably involved in the responses to extracellular signals such as the presence of nutrients, ions, or chemical messengers.

The genome of *Monosiga brevicollis* is estimated to encode as many as 128 tyrosine kinases, and approximately 380 total protein kinases [82] (For comparison, the human kinome contains 518 protein kinases, of which 90 are tyrosine kinases). Most of the *Monosiga brevicollis* tyrosine kinases have no identifiable metazoan homologues (the exceptions are nonreceptor tyrosine kinases of the Src, Csk, Abl, and Tec families). As in metazoans, *Monosiga brevicollis* tyrosine kinases never appear as isolated catalytic domains. Many of the domain combinations are distinct from any observed in metazoans. There are predicted to be 88 RTKs in *M. brevicollis*; 73 of the RTKs cluster into 15 families (designated RTKA, RTKB, etc.) [82]. While these RTKs possess many of the same features as metazoan RTKs (putative extracellular ligand-binding modules, conserved catalytic residues, and potential sites of autophosphorylation), their overall sequences show limited homology to the families of metazoan RTKs. It is not yet known whether the *M. brevicollis* RTKs (or any unicellular RTKs) are enzymatically active as tyrosine kinases.

The RTKB family of tyrosine kinases from *M. brevicollis* consists of nine members. While the sequences of the RTKB kinase catalytic domains are related, the extracellular regions are quite diverse. Three of the nine RTKB kinases appear to lack important catalytic residues and may function as pseudokinases or scaffolding proteins. Of the six remaining RTKB kinases, four possess a unique modular domain (designated RM2) in their cytoplasmic tails (Figure 5-1). The RM2 domains contain potential Src-family kinase phosphorylation sites and SH2-binding sites, suggesting that they may link RTK activation with downstream cytoplasmic signals [82]. In this chapter, we characterize *M. brevicollis* RTKB2, a kinase with six RM2 domains in its C-tail. We cloned, expressed, and purified the RTKB2 kinase, and conducted biochemical studies of its activity. This is the first investigation into the activity of an RTK from a unicellular organism.

S. rosetta STAT

Salpingocea rosetta is a choanoflagellate, closely related to *M. brevicollis* (Figures 1-5, 5-2). Its ability to live in both unicellular and colony forms makes it a good model to study signaling differences between the two life cycles. Investigations into its life history, cell differentiation and intercellular interaction have determined at least five distinct cell types *S. rosetta* can take in direct response to environmental cues: three solitary cell types (slow swimmers, fast swimmers, and thecate cells), and two colonial cell types (rosettes and chains) (Figure 5-3). These colonies form as a result of incomplete cytokinesis, rather than cell aggregation, so that the colony is composed of a cloned individual. Neighboring cells in colonies are connected by intercellular bridges that may allow the sharing of small molecules [91, 92]. 469 protein kinases have been identified in the *S. rosetta* genome; 376 of them are serine/threonine kinases which are highly conserved between *S. rosetta* and *M. brevicollis*. The tyrosine kinases are more divergent with large numbers of gains and losses, indicating rapid evolution. The nonreceptor tyrosine kinases are evolutionarily stable between *S. rosetta* and *M. brevicollis*, with 90% in common. In contrast with this, the receptor tyrosine kinases are more divergent, with only 21% orthologous between the two choanoflagellates. Interestingly, the *S. rosetta* tyrosine kinases that are more conserved with *M. brevicollis* are expressed more in attached cells, while the cells in rosette colonial form favor the tyrosine kinases unique to *S. rosetta* [118].

A STAT-like protein was identified in the genome of *S. rosetta*, with a similar domain architecture as mammalian STAT, and the conserved, functional residues are present (Figure 5-4). STAT DNA binding domains have also been found in the genomes of *Monosiga brevicollis* and *Capsaspora owczarzaki* [80]. In mammalian cells, the STAT proteins are a family of ligand-activated transcription factors that act both as signal transducers and activators of transcription (STAT) [119]. There are seven members of the STAT family: STAT1, STAT2, STAT3, STAT4, STAT5A, STAT5B, and STAT6. STATs are activated through phosphorylation by Jak-family kinases associated with growth factor receptors or cytokine receptors, or by other tyrosine kinases such as Src. The phosphorylated receptors (or nonreceptor tyrosine kinases) interact with STAT SH2 domains, allowing STAT phosphorylation. Once phosphorylated, STATs dimerize through reciprocal phosphotyrosine-SH2 domain interactions, allowing them to translocate into the nucleus and induce gene expression [120-122]. In this chapter, we investigate the ability of *S. rosetta* Src to phosphorylate *S. rosetta* STAT. The ability of the choanoflagellates to adopt simple forms of multicellularity, and the presence of tyrosine kinases in their genomes, makes them a good model to provide mechanistic insights into early stages in the evolution of animal multicellularity.

RTKB2 Results

Structure of RTKB2. *Monosiga brevicollis* RTKB2 is predicted to be a type I transmembrane protein, with a single membrane-spanning region [82] (Figure 5-1A). The extracellular domain organization of RTKB2 has several features in common with families of metazoan RTKs, although the exact combination of domains has not been observed in any metazoan family. RTKB2 contains a cystine-rich domain similar to those seen in the TNF receptor and to the furin-like domains of EGFR and insulin receptor [10]. Two divergent repeats similar to hyaline are present; these modules are structurally related to the immunoglobulin-like fold seen in many metazoan RTKs [123]. RTKB2 contains a short consensus repeat (SCR)/complement control protein domain, as seen in a variety of complement and adhesion proteins. Towards the C-

terminus of the extracellular domain, RTKB2 contains five EGF-like modules, as seen in (for example) the Tie and Eph families of metazoan RTKs. As is true for all *Monosiga brevicollis* RTKs, the extracellular ligand is unknown.

The predicted intracellular portion of RTKB2 contains a single tyrosine kinase catalytic domain. The kinase domain of RTKB2 has limited homology to known families of metazoan RTKs, but possesses the catalytically important residues found in all tyrosine kinases (Figure 5-1B). RTKB2 has a single tyrosine residue in the predicted activation loop. The C-terminus of RTKB2 contains 6 repeats of a unique domain designated RM2. RM2 domains are composed of approximately 80 amino acid residues. They are found in the cytoplasmic tails of four of the nine *Monosiga brevicollis* RTKB tyrosine kinases, but they are not present in any other organism [82]. RM2 domains contain tyrosine residues that are predicted to serve as phosphorylation sites as well as SH2-binding sites.

Characterization of RTKB2 Activity. (Experiments for this subsection were carried out by Dr. Todd Miller). To test whether the predicted RTKB2 protein is an active tyrosine kinase, we used PCR to amplify the RTKB2 catalytic domain (residues 1450-1724) from an *M. brevicollis* cDNA library. We cloned the catalytic domain into a baculovirus expression vector, infected Sf9 insect cells and purified the protein. Preliminary experiments showed that the purified RTKB2 was highly active, with a specific activity similar to other purified RTK kinase domains (such as the kinase domain of insulin-like growth factor I receptor) [124]. The highest activity was observed with a peptide (E4YM4) containing the EEEEYMMMM motif that was selected from a synthetic peptide library as a highly efficient insulin receptor substrate [125]. RTKB2 also had activity toward peptides derived from the fifth and sixth RM2 domains on RTKB2; these peptides can also be phosphorylated by the *Monosiga* MbSrc1.

Many RTKs are regulated by autophosphorylation at one or more tyrosines within the activation loop. RTKB2 has a single tyrosine in the predicted activation loop (Figure 5-1B). We tested for autophosphorylation by treating the purified enzymes with *Yersinia* tyrosine-specific YOP phosphatase and then incubating with magnesium and [γ^{32}P]ATP and following the reaction by SDS-PAGE and autoradiography. The stoichiometry of phosphorylation after 15 minutes was 0.93 mol phosphate/mol protein. Dephosphorylation of RTKB2 (by YOP) led to a loss of activity. The activity was regained upon autophosphorylation.

RM2 is a Substrate for a SFK. The six RM2 domains in the C-tail of RTKB2 contain tyrosines near their N-termini that are predicted to serve as phosphorylation and SH2-binding sites (Figure 5-5). As mentioned above, synthetic peptides derived from two of the domains are indeed substrates for the RTKB2 kinase. To study an intact RM2 domain, the gene corresponding to the sixth RM2 domain was synthesized. To determine whether the isolated RM2 domain could serve as a substrate for MbSrc1, we incubated the two purified proteins with [γ^{32}P]ATP and stopped reactions after 0, 2, 5, and 10 minutes. Mixtures were separated by SDS-PAGE and analyzed by autoradiography. Bands were excised and quantitated by scintillation counting. After 5 minutes, the stoichiometry was 2.02 ± 0.13 moles of phosphate per mole of RM2-6 domain (Figure 5-6). RM2-6 has three potential tyrosine phosphorylation sites; these results raise the possibility of multisite phosphorylation in the six RM2 domains in the C-terminus of RTKB2. The Tyr-Glu-Ala-Ile sequence within RM2-6 fits the consensus sequence for binding to the SH2 domains of Src-family kinases [125]. The MbSrc1 SH2 domain has a similar binding preference as its metazoan counterparts [93]. To test for MbSrc1 SH2 binding, we phosphorylated RM2-6 using RTKB2 and [γ-^{32}P]ATP. The pY-RM2 was incubated with an immobilized protein consisting of glutathione S-transferase (GST) fused to the SH3 domain of MbSrc1. The phosphorylated RM2 domain bound to the MbSrc1 SH2 domain (this experiment done by Dr. Todd Miller).

Expression of RM2 in Mammalian Cells. To study RM2 domain phosphorylation in a cellular context, we amplified the cDNA encoding the C-terminal tail of the RTK (all six RM2 domains) and cloned it into a mammalian expression vector to produce a fusion with green fluorescent protein (GFP). We expressed the protein in Src/Yes/Fyn triple knockout fibroblast cells (SYF), which lack all Src family kinases. Fluorescence microscopy showed that the GFP-RTKB2-CT has a diffuse cytoplasmic localization (Figure 5-7). We then expressed GFP-RTKB2-CT in SYF cells alone or in the presence of Flag-tagged MbSrc1 kinase (Figure 5-8). We isolated RTKB2 by anti-GFP immunoprecipitation and analyzed tyrosine phosphorylation by Western blotting. RTKB2-CT was weakly tyrosine-phosphorylated in these experiments, and the addition of MbSrc1 gave only a modest increase in phosphorylation (Figure 5-8). We were unable to detect a complex between RTKB2-CT and MbSrc1 by co-immunoprecipitation binding assays (data not shown). While these experiments confirm the ability of the RTKB2 C-tail to be phosphorylated, it is likely that ligand-stimulated RTKB2 kinase would produce much higher levels of tyrosine

phosphorylation. Furthermore, the lack of membrane-anchoring motifs in the RTKB2-CT and MbSrc1 constructs likely reduces the efficiency of their interaction.

S. rosetta STAT Results

Identification of *S. rosetta* STAT. A STAT-like protein was identified in the genome of *S. rosetta*. It has a similar domain architecture as its mammalian counterpart: an N-terminal STAT protein-protein interaction domain, a coiled-coiled domain, STAT binding domain, followed by an SH2 domain (Figure 5-4). *S. rosetta* STAT also displays high homology with key areas of mammalian STAT for proper functioning. These areas include amino acid residues that interact with DNA, amino acids in the SH2 domain that bind a phosphotyrosine region in the active homodimer and the corresponding residues in the phosphotyrosine region that are bound by the SH2 domain (Figure 5-4).

In mammalian cells, Src is able to phosphorylate STAT, thus promoting the activated dimer, which can then translocate to the nucleus to alter gene expression. We tested the ability of mammalian Src and *M. brevicollis* MbSrc1 to phosphorylate a *S. rosetta* STAT-derived peptide *in vitro* (peptide received from Dr. Nicole King and Tera Levin from the University of California at Berkeley). This peptide (GEAKNSKLLTKGYIPA) contains the tyrosine residue that aligns with the major phosphorylation site in mammalian STATs. We carried out enzyme reactions with [γ-^{32}P]ATP and analyzed the results using the phosphocellulose paper assay. Both enzymes were able to phosphorylate the *S. rosetta* STAT peptide. v-Src had lower activity toward the peptide than the more closely related MbSrc1 (Figure 5-9).

To investigate this relationship in mammalian cells, we co-expressed V5-tagged *S. rosetta* STAT in Src/Yes/Fyn-deficient fibroblast cells with either FLAG-tagged mammalian c-Src or *S. rosetta* Src. The cloned DNA constructs were given to us by Dr. Nicole King and Tera Levin, University of California at Berkeley). STAT was immunoprecipitated with anti-V5 antibody, and probed with antiphosphotyrosine antibody to visualize phosphorylation. When the *S. rosetta* STAT was expressed alone, no phosphorylation was seen. Co-expression of c-Src or

S. rosetta Src led to dramatic increases in the phosphorylation of *S. rosetta* STAT (Figure 5-10). The levels of phosphorylation were comparable for c-Src and *S. rosetta* Src.

Discussion

The abundance and diversity of receptor and nonreceptor tyrosine kinases in the unicellular choanoflagellate *Monosiga brevicollis* rival that of any metazoan [80-82]. The 88 RTKs found in the genome of *M. brevicollis* possess a wide variety of domain organizations. The divergent architectures of the choanoflagellate RTKs were likely generated by gene duplication and domain shuffling [126, 127]. The *M. brevicollis* RTKs have no direct homologues in multicellular organisms. In contrast, sponges, which are regarded as the oldest surviving metazoan lineage, possess most of the RTK families found in higher metazoans [53, 128, 129]. The genome of the sponge *Amphimedon queenslandica* contains 150 RTK genes, including kinase domains from six known animal families: epidermal growth factor receptor (EGFR), Met, discoidin domain receptor (DDR), ROR, Eph, and Sevenless [128]. The sponge *Oscarella carmela* possesses a similar array of homologs, and RTKs with homology to the receptors for insulin-like growth factor I and fibroblast growth factor [129]. This is consistent with a model in which the common ancestor between choanoflagellates and metazoans had RTKs, but the animal cell-specific families of RTKs developed after the split between the two groups [82].

RTKB2, the tyrosine kinase studied here, is one of the nine RTKB-family kinases from *M. brevicollis*. As the most primitive RTK to be studied yet, these results shed light on the evolution of biochemical function in the receptor tyrosine kinase superfamily. RTKB2 is active as a tyrosine kinase, and the intrinsic enzymatic function of the RTKB2 catalytic domain is high toward synthetic peptides, particularly the IR/IGF1R family peptide substrate E4YM4. RTKB2 also catalyzes autophosphorylation, an event that increases the activity of the enzyme. RTKB2 possesses a single tyrosine residue within the predicted activation loop. Activation loops are one

of the distinguishing features of eukaryotic protein kinases [130]. They are flexible, dynamic segments that are often stabilized in the active conformation by addition of one or more phosphates (either through autophosphorylation or through the action of another kinase). The control of RTKB2 activity by autophosphorylation indicates that this mode of regulation was present in primitive RTKs before the evolution of multicellular animals over 600 million years ago.

The C-terminal tail of RTKB2 contains 6 copies of the RM2 domain, a unique region of approximately 80 residues that has not been found outside of the RTKB family in *M. brevicollis* (Figure 5-1). RTKB1, RTKB3, and RTKB4 each have one copy of the RM2 domain C-terminal to their tyrosine kinase domains. Each of the RM2 domains has a tyrosine residue preceded by one or more negatively-charged amino acids. Scansite prediction indicated that these could serve as Src phosphorylation sites and/or SH2-binding sites [82]. Our data show that an isolated RM2 domain is phosphorylated by MbSrc1, a Src-family nonreceptor tyrosine kinase from *M. brevicollis* (Figure 5-6). The RTKB2 kinase domain also phosphorylates the isolated RM2 domain [131] (experiment done by Todd Miller). Small synthetic peptides derived from the potential phosphorylation sites are also substrates. The tyrosine phosphorylated RM2 domain binds specifically to the SH2 domain of MbSrc1, suggesting that phosphorylation of one or more RM2 domains can recruit cytoplasmic tyrosine kinases and other cellular proteins to propagate the RTKB2 signal [131] (experiment done by Todd Miller). In contrast to other modular signaling domains (eg., SH2 or SH3), the isolated RM2 domain appears to lack an ordered structure when it is removed from the context of the surrounding protein [131] (experiment done by Michael Tong and Markus Seeliger). These binding sites at the C-terminus or RTKB2 may serve a similar recruitment function as the tyrosine motifs found in the short, unstructured cytoplasmic tails of the thrombopoietin or erythropoietin receptors [132, 133] or the phosphorylation sites in the C-termini of RTKs such as the epidermal growth factor receptor.

Binding of the *M. brevicollis* MbSrc1 kinase to the phosphorylated RM2 domain raises the possibility that receptor and nonreceptor tyrosine kinase signaling are linked, as in metazoans. A classic example of the linkage in mammalian cells is observed in the role for Src in relaying the signal through the platelet-derived growth factor (PDGF) family of receptors. Src stably associates with the cytoplasmic portion of the PDGF receptor, leading to enhanced Src activity [134, 135] . MbSrc1 phosphorylates the C-tail of RTKB2 weakly in the absence of an

activating signal (Figure 5-8). Our results are consistent with a model in which RTKB2 activation (by an unknown signal) stimulates receptor autophosphorylation within the RM2 domains. This leads to MbSrc1 recruitment and further phosphorylation, which may serve to increase or dampen specific downstream signals. Identifying the nature of these signals will require the development of methodology to manipulate gene function in choanoflagellates.

A comparison of the 55 Mb *S. rosetta* genome with genomes from diverse opisthokonts such as *M. brevicollis, C. owczarzaki* and *M. vibrans*, suggests that the origin of metazoans was preceded by a period of high gene loss and gains. The *S. rosetta* genome encodes homologues of cell adhesion, neuropeptide, and glycosphingolipid metabolism genes previously found only in metazoans and increases the mounting list of genes believed to have been present in the progenitors of metazoans and choanoflagellates [118]. The identification of a STAT-like protein in the *S. rosetta* genome gave us another opportunity to study a unicellular Src family kinase substrate. Unlike many other cellular signaling cascades, the STAT pathway is direct: STATs bind to activated molecules at the cell membrane (growth factor or cytokine receptors, or nonreceptor tyrosine kinases such as Src) and translocate into the nucleus where they act as transcription factors. *S. rosetta* STAT possesses a similar domain arrangement as metazoan STAT: an N-domain responsible for dimer-dimer interactions, a coiled-coiled domain for protein interactions, a DNA-binding region, an SH2 domain for receptor binding and dimerization, followed by a site for tyrosine phosphorylation near the C-terminal [136] (Figure 5-4). *S. rosetta* STAT also contains conserved amino acid residues necessary for metazoan STAT functions such as binding to the DNA in the nucleus upon STAT activation, and interactions between the SH2 domain and phosphotyrosine region important for homodimerization (Figure 5-4).

In Metazoa, nonreceptor oncogenic tyrosine kinases such as v-Src are able to phosphorylate the STAT family members, particularly STAT3 and STAT5, promoting cancer cell growth [137, 138]. In these experiments, we tested whether *S. rosetta* STAT acted as a substrate for Src, as in metazoans. We found that the *S. rosetta* STAT serves as a substrate for *S. rosetta* Src or mammalian c-Src in SYF cells (Figure 5-10). Furthermore, a synthetic peptide derived from the putative *S. rosetta* STAT phosphorylation site was recognized by purified v-Src or *Monosiga brevicollis* MbSrc1 (Figure 5-9). This suggests that the mode of substrate recognition is conserved in choanoflagellates.

Our preliminary studies of two different Src family kinase substrates from choanoflagellates (the tail of the receptor tyrosine kinase RTKB2 of *M. brevicollis*, and a STAT from *S. rosetta)* give valuable insight into the evolution of signaling mechanisms in pre-metazoans. Both the ability of *S. rosetta* STAT to be phosphorylated by Src, and the indicated connection between receptor and nonreceptor tyrosine kinases in *M. brevicollis* shows a strikingly similar mode of cellular communication in unicellular choanoflagellates to homologous systems in multicellular organisms.

Figure 5-1. **Domain organization of the M. brevicollis RTKB2 kinase.** (A) Domain abbreviations: Cys, cysteine-rich region; HYR, repeats similar to hyaline domains; SCR, short consensus repeat/complement control protein domain; E, epidermal growth factor repeat; tm, transmembrane sequence; RM2, unique M. brevicollis domain. (B) Amino acid sequence of the TRKB2 kinase domain, aligned with human fibroblast growth factor receptor 1 (FGFR1) and epidermal growth factor receptor (EGFR). Sequences were aligned using the ClustalW program and formatted with BOXSHADE. The position of the kinase-conserved DFG motif at the beginning of the activation loop is indicated with a red bracket, and an asterisk indicates the position of the potential autophosphorylation site.

A

B

```
RTKB2   1450  VRLGEELGHGAFGEVFQAWIR---NSYSGKVRRCAAKTLKPDVSFEHRQDFLSEMTVMKE
FGFR1   476   LVLGKPLGEGCFGQVVLAEAIGLDKDKPNRVTKVAVKMLKSDATEKDLSDLISEMEMMKM
EGFR    712   FKKIKVLGSGAFGTVYKGLWIP---EGEKVKIPVAIKELREATSPKANKEILDEAYVMAS

RTKB2   1507  IGAHPNVIGIIGHCLRETPQILLVELAEFGNMRDYLRGCR---------ATQENPQTIRV
FGFR1   536   IGKHKNIINLLGACTQDGPLYVIVEYASKGNLREYLQARRPPGLEYCYNPSHNPEEQLSS
EGFR    769   VDN-PHVCRLLGICLTST-VQLITQLMPFGCLLDYVR--------------EHKDNIGS

RTKB2   1558  DQMADFCLQIARGMAFLEQKNVIHRDLAARNVLVDKNFECKISDFGLARSIEGGEYKTTA
FGFR1   596   KDLVSCAYQVARGMEYLASKKCIHRDLAARNVLVTEDNVMKIADFGLARDIHHIDYYKKT
EGFR    812   QYLLNWCVQIAKGMNYLEDRRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEYHA

RTKB2   1618  T--KLPVKWMAPESLKSRIYTTKSDVWSFGVTMWEIFSLGGTPYKEHQNRDVLSFLESGH
FGFR1   656   TNGRLPVKWMAPEALFDRIYTHQSDVWSFGVLLWEIFTLGGSPYPGVPVEELFKLLKEGH
EGFR    872   EGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSILEKGE

RTKB2   1676  RLPSPRQAPVEFDDIARMCWVLEASDRPSFSELVDLLADIVL
FGFR1   716   RMDKPSNCTNELYMMMRDCWHAVPSQRPTFKQLVEDLDRIVA
EGFR    932   RLPQPPICTIDVYMIMVKCWMIDADSRPKFRELIIEF-----
```

Figure 5-2. ***Salpingocea rosetta: a choanoflagellate.*** Image of *S. rosetta.*

Picture from Dr. Nicole King, University of California at Berkeley

Figure 5-3. **A model of *S. rosetta* life history.** *S. rosetta* cells can differentiate into at least 5 different forms: fast swimmers, slow swimmers, thecate cells, chain colonies, and rosette colonies. Arrows depict observed transitions between cell types.

Figure from Dayel et al, Dev. Biol., **357**, 73-82 (2011).

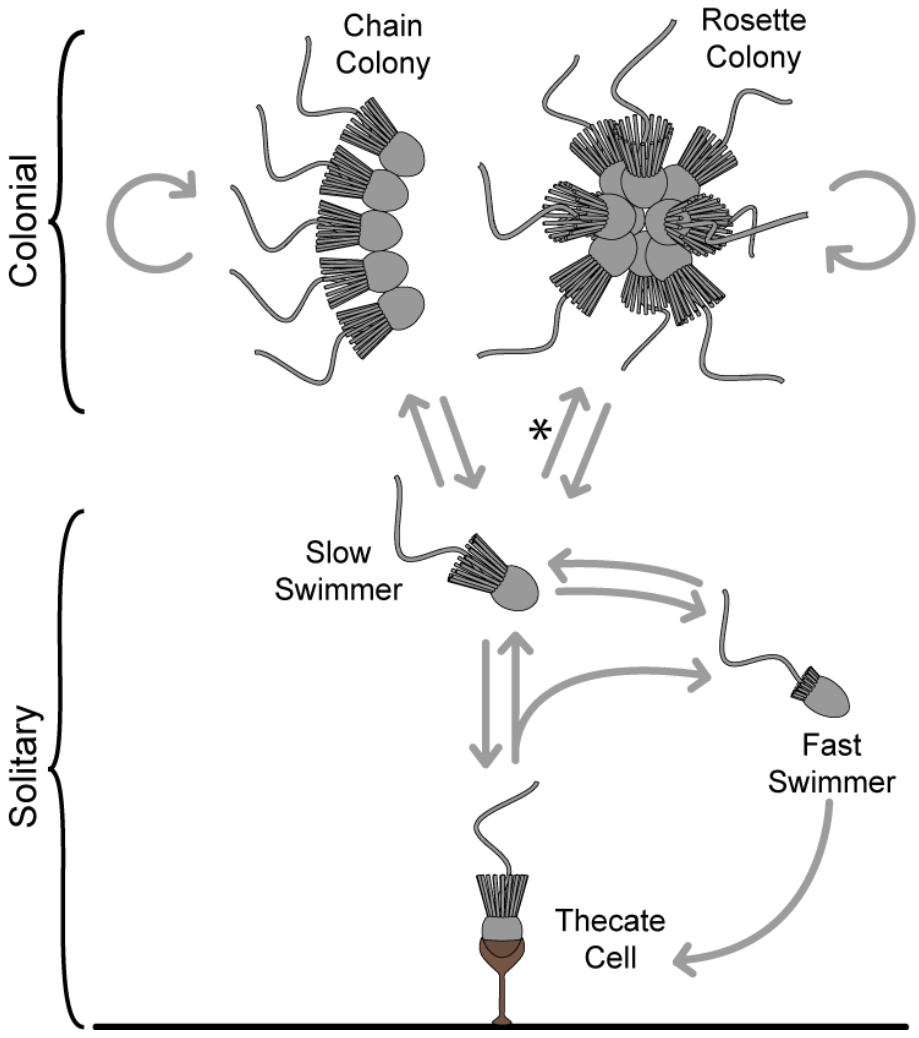
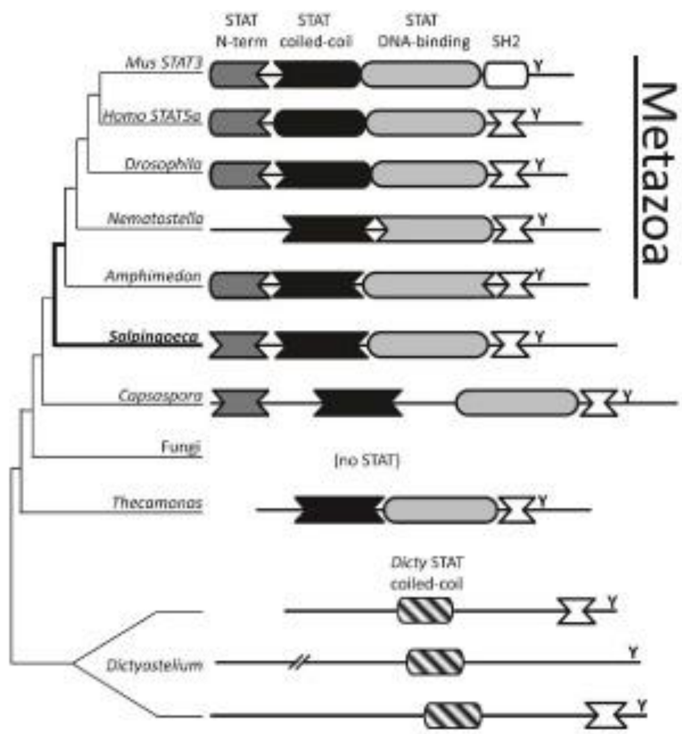

Chain
Colony
Rosette
Colony
Colonial
*
Slow
Swimmer
Fast
Swimmer
Solitary
Thecate
Cell

Figure 5-4. **S. rosetta STAT has conserved domain arrangement and critical residues from mammalian STAT.** (A) STAT domains (N-terminal STAT protein-protein interaction domain, DNA-binding domain, SH2 domain, and Dictyostelium coiled-coil domain) were predicted using Pfam. Round ends represent regions where domain was entirely predicted, the jagged where partial domain was predicted. The Y shows the location of the critical tyrosine residue that is phosphorylated upon STAT activation. (B) Regions of multiple sequence alignment between S. rosetta and metazoan STATs. Residues are highlighted based on identity/similarity in four or more species. Gray circles represent amino acids that bind DNA, open circles represent SH2 domain amino acids that interact with phosphotyrosine region in the active homodimer, black circles represent amino acids in the phosphotyrosine region that interact with the SH2 domain in the active homodimer.

Figure from Tera Levin and Nicole King (University of California at Berkeley)

A



B

Salpingoeca	-------PQVLKTQSRFQA	-------	RKIKRTD	DHTVTEQ	SLPVVVVVHGNQVPGAEA
Amphimedon	-------PQVIKTSNRFSA	-------	RRIKRNS	TEIVMEE	SLPVVVIVHVTQQPAAEA
Nematostella	-------PQILKIGTKFCA	-------	TKTDREG	DETVCEK	SFPCVVTVHGNQSPDAEA
Drosophila	-------PQVMKTNTRFAA	-------	KKIKRAE	TESVMDE	SLPVVVIVHGNQEPQSWA
Mus STAT3	PMHPDR-PLVIKTGVQFTT	RGSRKFN	REQRCGN	SLIVTEE	SLPVVVISNICQMPNAWA
Homo STAT5a	-------PQVLKTQTKFAA	KNENTRN	KRIKRAD	AESVTEE	SLPVVVIVHGSQDHNATA

Salpingoeca	FIDKDET	TFLVRFSTS-LVS
Amphimedon	FMDKIEA	TFLVRFSDS-EAG
Nematostella	FLSKQQA	TFLLRFSDN-ELG
Drosophila	FINKTKA	TFLLRFSDS-ELG
Mus STAT3	FISKERE	TFLLRFSESSKEG
Homo STAT5a	FVNKQQA	TFLLRFSDS-EIG

Salpingoeca	GYIPADL
Amphimedon	GYLESQV
Nematostella	EYRPTQL
Drosophila	GYVKSTL
Mus STAT3	PYLKTKF
Homo STAT5a	GYVKPQI

Figure 5-5. **Alignment of the amino acid sequences of the six RM2 domains.** Sequences were aligned using the ClustalW program and formatted with BOXSHADE. The conserved tyrosine residue is indicated by a red asterisk. The sixth RM2 domain, shown in red, was chosen for further analysis by bacterial expression.


RM2-1
RM2-2
RM2-3
RM2-4
RM2-5
RM2-6

Figure 5-6. **Phosphorylation of RM2 domain by MbSrc1.** Purified RM2-6 was phosphorylated by MbSrc1 in the presence of [γ-^{32}P]ATP . Samples were removed at 0, 2, 5 and 10 minute intervals, and reactions stopped by addition of Laemmlie buffer. Mixtures were separated by SDS-PAGE. Bands were excised and analyzed by Scintillation counting.

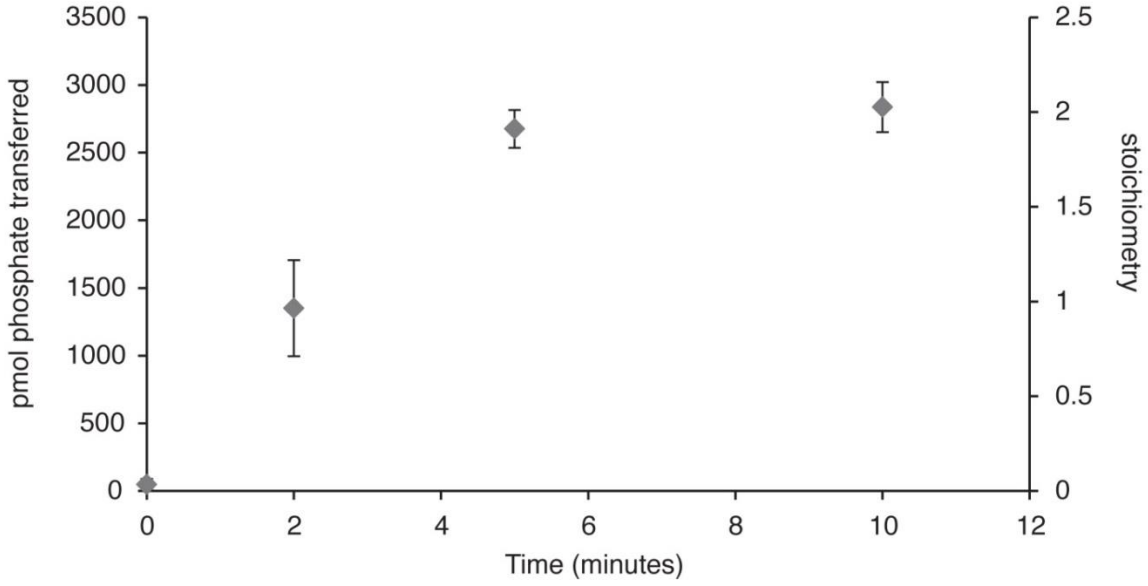
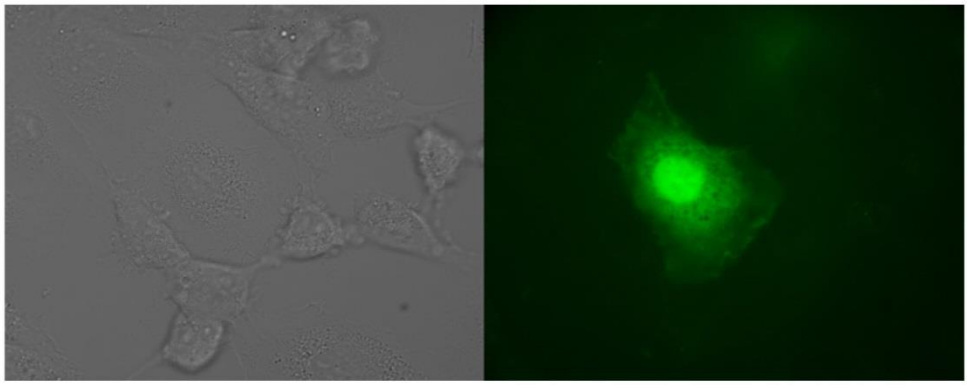

pmol phosphate transferred
3500
3000
2500
2000
1500
1000
500
0
2.5
2
1.5
1
0.5
0
stoichiometry
0
2
4
6
8
10
12
Time (minutes)

Figure 5-7. **GFP-RTKB2-CT has a diffuse cytoplasmic localization.** Immunofluorescence microscopy of SYF cells expressing GFP (top) or GFP-RTKBC-cyto (bottom)

A

GFP



B

GFP-RTKB2-cyto

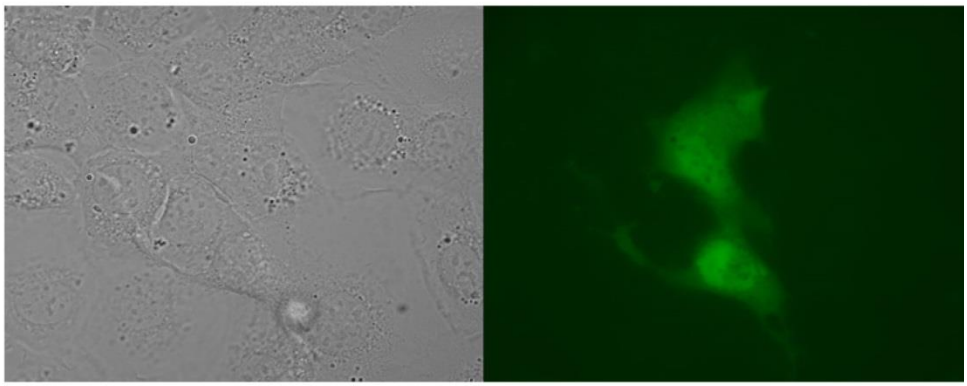


Figure 5-8. **Expression and Phosphorylation of the C-terminus of RTKB2 in mammalian cells.** SYF cells were transiently transfected with plasmids encoding GFP or GFP-RTKB2-C-terminus (alone or together with a plasmid encoding FLAG-tagged MbSrc1). Top panels: anti-GFP immunoprecipitates were separated by SDS-PAGE and analyzed by Western blotting with anti-phosphotyrosine antibodies. The membrane was subsequently stripped and reprobed with anti-GFP antibodies. The bottom two panels show expression of GFP, GFP-RTKB2-CT, and FLAG-MbSrc1 in the SYF cell lysates.

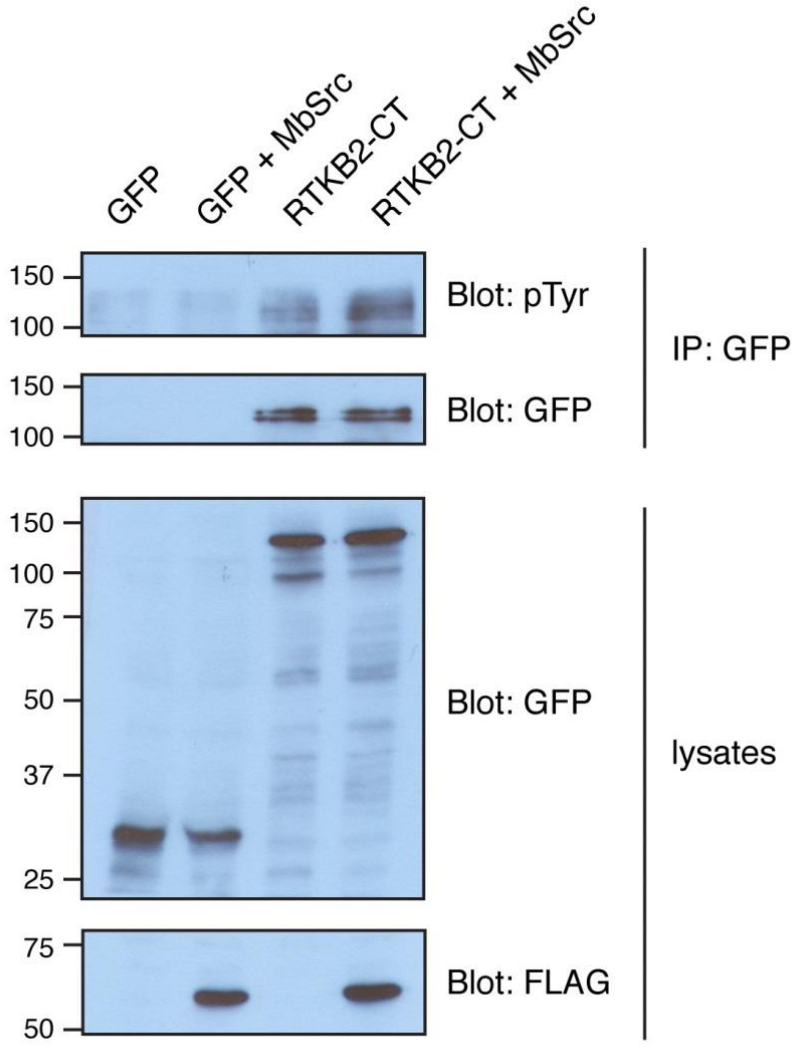

GFP
GFP + MbSrc
RTKB2-CT
RTKB2-CT + MbSrc
150
100
Blot: pTyr
150
100
Blot: GFP
IP: GFP
150
100
75
50
37
25
Blot: GFP
lysates
75
50
Blot: FLAG

Figure 5-9. ***S. rosetta* STAT is a Src substrate.** Activity of mammalian v-Src and *Monosiga brevicollis* MbSrc1 (1.5 μM) were assayed with 750 μM of Src peptide (AEEEIYGEFEAKKKKG) or *S. rosetta* STAT peptide (GEAKNSKLLTKGYIPA). Enzyme activities were measured for 20 minutes with 0.25 mM [γ-^{32}P]ATP using the phosphocellulose paper binding assay.

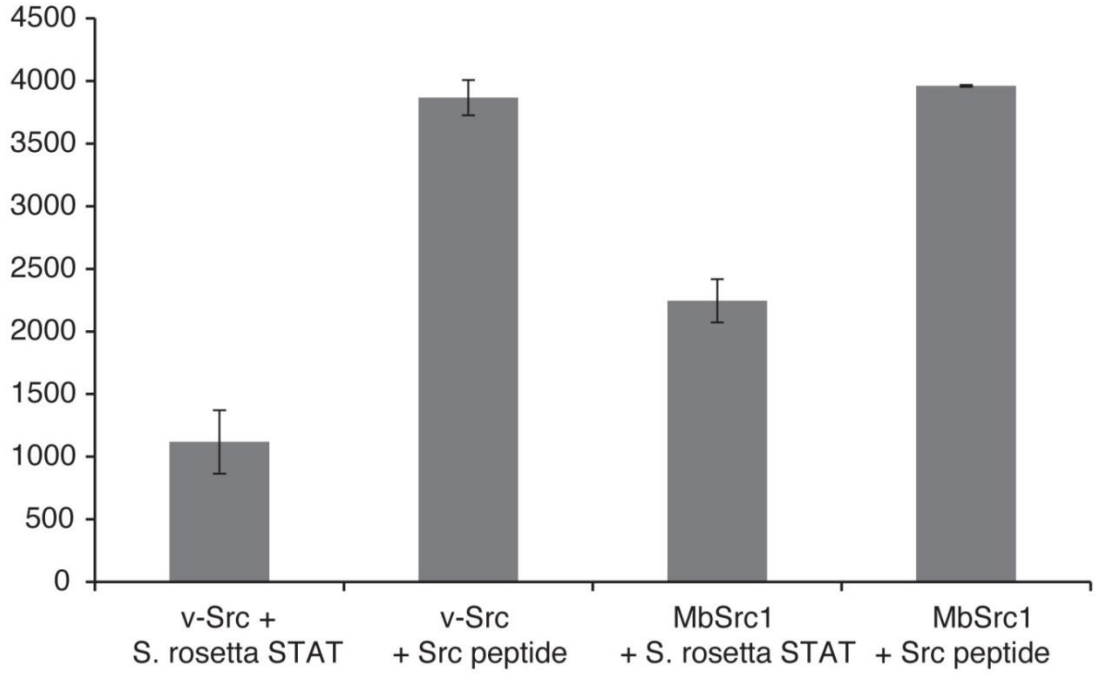
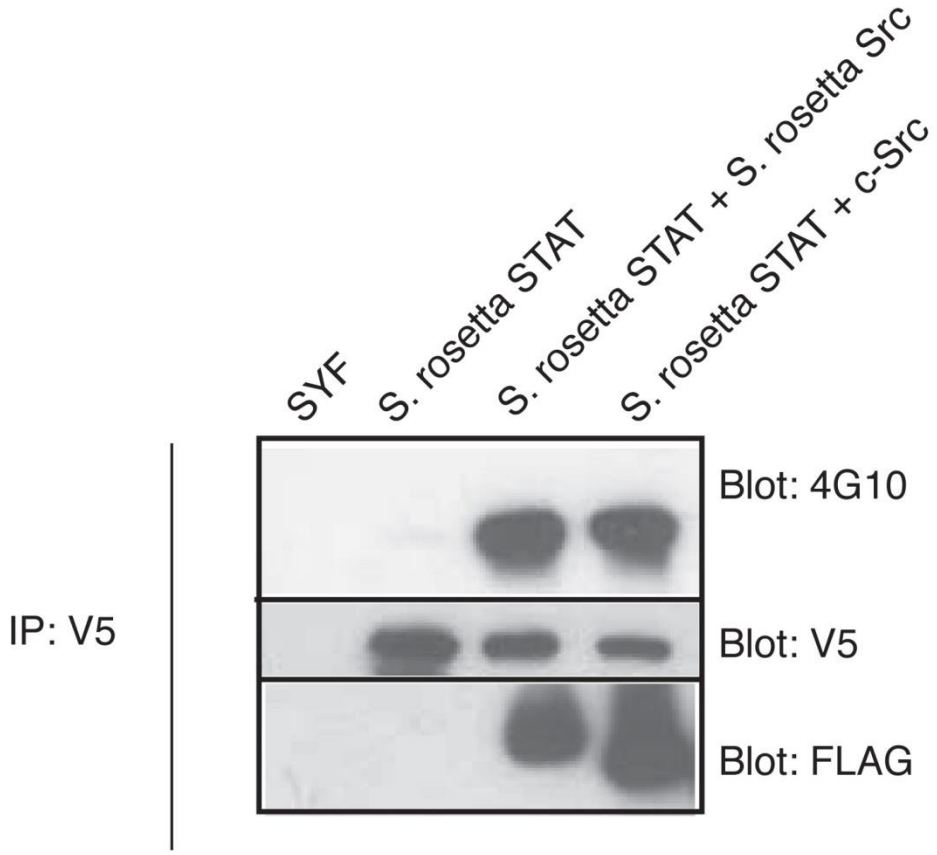

4500
4000
3500
3000
2500
2000
1500
1000
500
0
v-Src +
S. rosetta STAT
v-Src
+ Src peptide
MbSrc1
+ S. rosetta STAT
MbSrc1
+ Src peptide

Figure 5-10. **Phosphorylation of *S. rosetta* STAT by Src in mammalian cells.** SYF cells were transiently transfected with V5-tagged *S. rosetta* STAT alone, or in the presence of FLAG-tagged *S. rosetta* Src or mammalian c-Src. STAT was immunoprecipitated by immobilized anti-V5 antibody. Pull-downs were probed with antiphosphotyrosine antibody, as well as anti-V5 antibody and anti-FLAG antibody for expression.


SYF
S. rosetta STAT
S. rosetta STAT + S. rosetta Src
S. rosetta STAT + c-Src
IP: V5
Blot: 4G10
Blot: V5
Blot: FLAG

Chapter 6

Concluding Discussion and Future Directions

The experiments in Chapter 3 build on previous studies of Src regulation in unicellular choanoflagellates. *Capsaspora owczarzaki* was chosen as the focus of our work because it represents the closet known relative to both the choanoflagellates and Metazoa. The genome of *Capsaspora owczarzaki* contains103 protein tyrosine kinase genes, including Src and Csk homologues. The two *C. owczarzaki* Src-like kinases, CoSrc1 and CoSrc2, both possess the expected domain architecture: SH3, SH2, and kinase catalytic domains (Figure 3-1). They also contain the conserved tyrosines important for Src regulation (corresponding to Tyr416 in the activation loop and Tyr 527 in the C-terminal tail). CoSrc1 and CoSrc2 display a similar substrate specificity as both mammalian c-Src and the choanoflagellate *M. brevicollis* MbSrc1 (Figure 3-2), and are able to functionally replace mammalian Src in SYF cells. A luciferase reporter assay showed that CoSrc2 activity was even higher than c-Src activity in SYF cells. This may be due to a lower level of autoregulation of the *Capsaspora* Src kinases in cells. The reliance of CoSrc activity on autophosphorylation in the activation loop is also shared with MbSrc1, and is indeed a characteristic of many tyrosine kinases. There is a lack of SH2-mediated substrate targeting in both CoSrc enzymes, a feature shared with MbSrc1. The absence of functional SH2-kinase domain coupling in both choanoflagellates and *Capsaspora* indicates that this is a feature of Src that developed later in evolutionary time, rather than a trait specifically lost in the choanoflagellate clade.

The localization of CoSrc1 and CoSrc2 into punctate structures in the filopodia of *Capsaspora*, combined with the revelation that its genome contains a complete set of integrin-mediated machinery, indicates to us the presence of primitive focal adhesion complexes. *Capsaspora* filopodia are used to crawl along a substrate. We hypothesize that *Capsaspora* cells have focal adhesion complexes in order to generate force to move across a surface. It is likely that early metazoans co-opted this machinery to create the modern focal adhesion system, allowing communication between cells and the extracellular matrix, rather than the *Capsaspora* cells with their environment.

A striking difference we observed between *Capsaspora* Csk and the previously studied Csk kinases from choanoflagellates (*Monosiga brevicollis* and *Monosiga ovata)*, is that CoCsk has no observable enzymatic activity. Not only is it unable to phosphorylate a synthetic general tyrosine kinase substrate, it is unable to phosphorylate purified CoSrc (Figures 3-9, 10, 11).

Further studies show that CoCsk does not impose negative regulation upon CoSrc activity (Figures 3-13, -14). Indeed, our results in mammalian cell lysates indicate an increase in activity when CoSrc kinases are co-expressed with CoCsk (Figure 3-13B). Since we showed a binding capability between CoSrc and CoCsk (Figure 3-12A), this may be a result of a partial release of autoinhibition due to the interaction of CoSrc and CoCsk. In *M. brevicollis*, MbCsk was active toward MbSrc1, but this phosphorylation did not reduce MbSrc activity [93]. This stands in contrast to studies from *E. fluviatilis*, a sponge representing an ancient metazoan, in which Csk inhibited Src [87]. These studies indicate a trend, in which the farther we move back in evolutionary time from the metazoan lineage, the less of a role Csk plays in Src activity.

Our work in *Capsaspora* cells supports our theory of low Src negative regulation. We saw comparable levels of CoSrc activity in *Capsaspora* lysates as in SYF cells over-expressing CoSrc and CoCsk (Figure 3-16A, B). The levels of endogenous CoSrc were undetectable by Western blotting as compared to levels in SYF lysates (Figure 3-16C). This suggests that endogenous CoSrc kinases are very active. This is similar to results seen in *M. brevicollis* cells, where MbSrc1 was unregulated by MbCsk and activity of MbSrc1 was therefore very high [93]. To determine the specific activity of endogenous CoSrc kinases, we need a better way to quantify the amounts of protein in *Capsaspora* cells. Testing the physiological roles of CoSrc and CoCsk in intact *Capsaspora* cells will hopefully be possible in the future as technology is developed to introduce genes or block gene function in the organism.

In Chapter 4, we characterize Src and Csk from *Ministeria vibrans.* *Ministeria* represents the other known organism to form the Filasterea clade with *Capsaspora.* By studying the signaling molecules in *Ministeria*, we are able to determine whether the lack of Csk-mediated Src in *Capsaspora* is a clade-wide trend, or whether this surprising result is an organism-specific phenomenon. We characterized two Src-like kinases and one Csk homolog in *Ministeria vibrans.* Similarly to *Capsaspora* CoSrc1 and CoSrc2, MvSrc and MvSrc2 displayed the same domain arrangement conserved in mammalians (Figure 4-1). MvSrc1 and MvSrc2 also preferred a Src substrate peptide, and showed the same lack of SH2 substrate targeting (Figures 4-3, 4-5). However, MvSrc2 also proved capable of phosphorylating a Ser/Thr kinase substrate at a serine residue (Figures 4-3, 4-4). Normally, purified tyrosine kinases are extremely specific for tyrosine, and will not phosphorylate serine or threonine, even when they are substituted for tyrosine in a substrate sequence, nor do serine/threonine kinases usually phosphorylate tyrosine

[139]. However, dual-specificity kinases, which are more ancient than tyrosine kinases, can carry out low levels of tyrosine phosphorylation. This is highlighted by the fact that phosphoamino acid analysis of yeast (which lack tyrosine kinases), have small populations of phosphotyrosine, the results of these dual-specificity kinases [79]. One example of this group of proteins is the MAP kinase kinases, which need to phosphorylate the MAPKs on both a threonine and a tyrosine residue within its activation loop. Sequence analysis has shown that tyrosine kinases and serine/threonine kinases belong to the same protein kinase superfamily. They have a number of residues in common that are highly conserved, but are distinguished by specific signature motifs [140]. The difference between substrate recognition of serine/threonine kinases and tyrosine kinases lies in the structural conformation: the backbone needs to be positioned differently in the two kinases in order to accommodate either aliphatic or aromatic hydroxyls on the serine/threonine or tyrosine kinase, respectively. It is hypothesized that tyrosine kinases evolved as a branch of the older serine/threonine kinases. Nine kinases in the *M. brevicollis* genome have dual catalytic domains, though in all cases one of the two domains is predicted to be catalytically inactive [82]. MvSrc2 represents a unique kinase, which is classified as a tyrosine kinase by sequence, yet is able to phosphorylate a serine/threonine kinase substrate. If tyrosine kinases evolved from serine/threonine kinases, as we hypothesize, MvSrc2 may be an example of an intermediary protein, newly evolved from serine/threonine kinases.

Functional assays with *M. vibrans* MvCsk proved it to be catalytically dead, similar to results with CoCsk in Chapter 3. MvCsk is unable to phosphorylate a general tyrosine kinase substrate (Figure 4-9), or purified MvSrc (Figure 4-10). MvCsk does not inhibit MvSrc1 or MvSrc2 *in vitro* (Figure 4-12). We observed a slight increase in MvSrc activity in the presence of MvCsk. As hypothesized for CoSrc, this may be due to an interaction between MvSrc and MvCsk that releases any autoinhibition present in the purified kinase. Binding assays between MvSrc and MvCsk would need to be performed to test for an interaction. The complete lack of phosphorylation or inhibition of MvSrc by MvCsk supports our hypothesis from Chapter 3, that the uncoupling of Src and Csk as a regulatory system is a distinctive feature of the filasterean lineage. There are two possibilities that may exist to explain these results: Csk evolved as an inactive kinase and gradually evolved into a regulator of Src, or alternatively, it emerged as an active enzyme, but lost its activity in the filasterean lineage. If Csk first developed as an inactive kinase, it is possible that the unicellular Csk counterparts could be classified as pseudokinases.

Although by sequence CoCsk and MvCsk would not be classified as pseudokinases (they both contain the three critical motifs in the catalytic domain: VAIK, HRD, and DFG), it is possible they play a role in other common pseudokinase activities such as scaffolding, or activating other proteins [70, 71]. Some pseudokinases, such as the WNK kinase, have been shown to be enzymatically active through alternative mechanisms [72]. It is possible that MvCsk and CoCsk have activities that are distinct from their mammalian counterparts, and of which we are unaware. To resolve these various possibilities, more studies are required with organisms from other, more evolutionarily ancient clades. The next step would be to analyze the genome(s) from the common ancestor of both Filasterea and chonaoflagellates, determine whether they express Src and Csk, and if so, characterize their activity. The Ichthyosporea seems to be the lineage that emerged just prior to Filasterea, and therefore organisms in this clade would be a good target for genomic analysis [141]. The idea that Csk began as an inactive pseudokinase, and slowly adopted activity as it evolved, has interesting implications in the way we view other pseudokinases. It is possible that some of these proteins arose with no enzymatic activity, but through evolutionary processes, have been co-opted to participate in complex systems. In mammalian cells, Csk enzymes play key roles in such diverse systems as cell proliferation, survival, and adhesion. Csk knock-out experiments that result in embryonic death highlight the medical importance of proper Src regulation. Experiments on the interaction of unicellular Src and Csk kinases give insight into the evolution of this complex relationship, as well as the possibility of other non-enzymatic functions of Csk. The tyrosine kinases encoded by proto-oncogenes in mammalian cells contain elaborate regulatory mechanisms that keep kinase activity in check. Mutations within these regulatory systems can result in constitutive kinase activation, loss of cell growth control, and cancer initiation and progression. An outstanding question is how the regulatory systems of these tyrosine kinases arose during evolution. Evolution can lead to the modification of existing TKs during the development of more complex signaling systems (e.g., the adaptive immune system). Understanding the molecular basis of this development can shed light on regions and residues of present-day proto-oncogenic kinases that are involved in allosteric regulation.

In Chapter 5, we discuss preliminary studies of Src substrates in choanoflagellates. The tail of RTKB2 contains 6 repeats of the RM2 domain. We found that this domain can serve as a substrate for a Src family kinase. When phosphorylated, the isolated RM2 domain binds to the

SH2 domain of MbSrc1 [131]. These results indicate a link between the receptor tyrosine kinases in choanoflagellates (which presumably interact with signals extracellular to the cell) and nonreceptor tyrosine kinases. In this way, signals can be transduced and propagated throughout the cell, as they are in Metazoa. In the same chapter, we discuss studies on another Src substrate in choanoflagellates, a STAT protein in *S. rosetta*. *S. rosetta* STAT contains all of the conserved residues necessary for mammalian STAT activity, including those involved in DNA interaction and STAT dimerization. We showed that *S. rosetta* STAT can be phosphorylated both *in vitro* and in mammalian SYF cells by mammalian and choanoflagellate Src kinases. This is another example of a cell-wide signaling system in unicellular choanoflagellates. Presumably, phosphorylation of the *S. rosetta* STAT by its Src protein would trigger dimerization and activation, allowing it to translocate to the nucleus and trigger gene expression. Further functional studies will need to be done to prove that *S. rosetta* STAT functions in a similar manner to its mammalian counterpart. Specifically, luciferase reporter assays would give insight into whether *S. rosetta* STAT has the same nuclear targets and functions as metazoan STAT.

The domain architectures seen in the choanoflagellate and filasterean lineages are more diverse than those seen in complex metazoans [82, 98]. The repertoire of tyrosine kinases in *M. vibrans* and *C. owczarzaki* suggests that tyrosine kinases evolved before the divergence of filasterea and the Metazoa and Choanoflagellate clades. The large numbers of tyrosine kinases with domain combinations not seen in Metazoa points to the theory that domain shuffling led to the architecture of mammalian Src family kinases. The modular properties of SFK domains make them evolutionarily flexible. By a method of "mixing and matching" functional domains, it is possible to produce a large number of diverse proteins with wide varieties of substrate recognition and targeting. Indeed, the choanoflagellate *M. brevicollis* genome contains 128 tyrosine kinases, while humans only express 90. It is possible that this domain shuffling "tested out" different domain architectures, and only the most successful were retained by multicellular animals [82]. In this case, a successful architecture could be viewed as one that fulfills the main requirements of metazoan tyrosine kinases: (1) the protein phosphorylates specific downstream proteins, and (2) the protein can be autoinhibited by intermolecular interactions.

The SH3-SH2-kinase domain arrangement, as well as being shared by the Src family kinases, is present in multiple nonreceptor tyrosine kinase families (Abl, Csk, Frk, and Tec families all share the same order of domains, although some of them have additional conserved

modular domains). It is important to note that when this combination of domains is present, this particular order is always preserved across all families. The same domain arrangement is conserved across all of the nonreceptor tyrosine kinase families in *Capsaspora owczarzaki* and *Ministeria vibrans* that contain all three domains. When Src family kinase mutants are produced with different orders of SH3, SH2, and kinase domains, changes in substrate recognition were seen (in addition to the changes in autoregulation that would be expected by such a vast change in enzyme structure) [35]. Although the strict requirement for this domain arrangement has been observed in the unicellular ancestors to Metazoa, it is unclear whether the cytoplasmic tyrosine kinase families adopted this architecture through convergent or divergent evolution. Either way, there is a clear evolutionary pressure to maintain this this domain order. Since both choanoflagellate and filasterean Src lacks autoregulation due to Csk phosphorylation, presumably the noncatalytic domains function in a substrate targeting capacity. Although we were only able to observe SH3 substrate targeting (the filasterean Src kinases recognized peptides containing an SH2 ligand or control equally well; Figures 3-3C and 4-5), a limitation of these experiments is that they were done using a peptide derived from the mammalian Src substrate. It is possible that endogenous protein substrates for unicellular Src kinases might still require SH2 binding. For example, the spacing between the SH2 ligand and substrate sequence might be different than that observed in c-Src. Experiments on mutant Src kinases with re-arranged domains showed that the placement of the SH3 domain is more flexible than the SH2 location [35]. This less-stringent requirement of the SH3 domain placement may explain why SH3 targeting was observed in our experiments, as opposed to SH2 targeting. Identifying the substrates of CoSrc and MvSrc from their respective genomes would allow us to more closely mimic the interactions that occur within the organism, and determine whether a more complex system of substrate specificity exists in filasterean cells than could be observed in our in vitro experiments.

The results of this thesis indicate that the unicellular ancestors in the metazoan lineage already had complex signaling systems in place. Because of current technical limitations, we were unable to determine what functional roles this machinery plays in the current organisms. We hypothesize that these signaling systems exist to communicate with the outside environment, and were later co-opted by multicellular animals. One indication of this role can be seen in the choanoflagellate *S. rosetta*, an organism capable of forming colonies. Neighboring cells in

colonies are connected by intercellular bridges, that may allow sharing of small molecules [92]. We can view this arrangement as a very primitive form of multicellularity, where signaling molecules can be passed through this loose "extracellular matrix." The development of multicellularity can therefore be considered as the transition from unicellular, separate cells, to an intermediate colony-forming step, which eventually became the complex, multicellular organisms seen in Metazoa.

It is important to note that this thesis focuses specifically on the emergence of multicellular organisms in the metazoan lineage. However, multicellularity appears in at least 16 diverse, independent eukaryotic lineages [142]. There is no reason to believe that there is one conserved pattern of multicellular emergence, especially due to the fact that Metazoa are the only lineage that contains tyrosine kinases. Advances in genomics research are revealing that across lineages, more of the genes associated with complex processes are already present in the simpler, unicellular ancestors [143]. This suggests that the presence of the genes cannot alone account for the evolution of multicellular organisms. Each time this transition occurred, signaling or adhesion mechanisms already in existence were co-opted for new use in the next, more complex form. By studying these signaling molecules, as well as identifying which genes are turned on and off in single cell vs. colonial forms, we can shed more light on one of the most pivotal (and least understood) events in metazoan history: the transition to multicellularity.

1. Hunter, T., *Tyrosine phosphorylation: thirty years and counting.* Current Opinion in Cell Biology, 2009. **21**(2): p. 140-146.
2. Manning, G., et al., *Evolution of protein kinase signaling from yeast to man.* Trends in Biochemical Sciences, 2002. **27**(10): p. 514-520.
3. Robinson DR, W.Y., Lin SF, *The protein tyrosine kinase family of the human genome.* Oncogene, 2000. **19**(49): p. 5548-57.
4. Brown, M.T. and J.A. Cooper, *Regulation, substrates and functions of src.* Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, 1996. **1287**(2–3): p. 121-149.
5. Lemmon, M.A. and J. Schlessinger, *Cell Signaling by Receptor Tyrosine Kinases.* Cell, 2010. **141**(7): p. 1117-1134.
6. Cadena, D.L. and G.N. Gill, *Receptor tyrosine kinases.* The FASEB Journal, 1992. **6**(6): p. 2332-7.
7. Koch, C.A., et al., *SH2 and SH3 Domains: Elements That Control Interactions of Cytoplasmic Signaling Proteins.* Science, 1991. **252**(5006): p. 668-674.
8. Blume-Jensen, P. and T. Hunter, *Oncogenic kinase signalling.* Nature, 2001. **411**(6835): p. 355-65.
9. Hubbard, S.R. and W.T. Miller, *Receptor tyrosine kinases: mechanisms of activation and signaling.* Current Opinion in Cell Biology, 2007. **19**(2): p. 117-123.
10. Hubbard, S.R. and J.H. Till, *PROTEIN TYROSINE KINASE STRUCTURE AND FUNCTION.* Annual Review of Biochemistry, 2000. **69**(1): p. 373-398.
11. Parsons, S.J. and J.T. Parsons, *Src family kinases, key regulators of signal transduction.* Oncogene, 2004. **23**(48): p. 7906-9.
12. Cole, P.A., et al., *Protein tyrosine kinases Src and Csk: a tail's tale.* Current Opinion in Chemical Biology, 2003. **7**(5): p. 580-585.
13. Irby, R. and T. Yeatman, *Role of Src expression and activation in human cancer.* Oncogene, 2000. **19**(49): p. 5636-5642.
14. Bromann PA, K.H., Courtneidge SA, *The interplay between Src family kinases and receptor tyrosine kinases.* Oncogene, 2004. **23**: p. 7957-68.
15. Pellman, D., et al., *An N-terminal peptide from p60src can direct myristylation and plasma membrane localization when fused to heterologous proteins.* Nature, 1985. **314**: p. 374-7.
16. Kaplan, J.M., et al., *The first seven amino acids encoded by the v-src oncogene act as a myristylation signal: lysine 7 is a critical determinant.* Molecular and Cellular Biology, 1988. **8**(6): p. 2435-2441.
17. Patwardhan, P., Resh, M, *Myristoylation and Membrane Binding Regulate c-Src Stability and Kinase Activity.* Molecular and Cellular Biology, 2010. **30**(17): p. 4094-4107.
18. Rudd, C.E., et al., *Pillars Article: The CD4 Receptor Is Complexed in Detergent Lysates to a Protein-Tyrosine Kinase (Pp58) from Human T Lymphocytes. Proc. Natl. Acad. Sci. USA 1988. 85: 5190–5194.* The Journal of Immunology, 2010. **185**(5): p. 2645-2649.
19. Shaw, A.S., et al., *The Ick tyrosine protein kinase interacts with the cytoplasmic tail of the CD4 glycoprotein through its unique amino-terminal domain.* Cell, 1989. **59**(4): p. 627-636.
20. Timson Gauen, L.K., et al., *p59fyn tyrosine kinase associates with multiple T-cell receptor subunits through its unique amino-terminal domain.* Molecular and Cellular Biology, 1992. **12**(12): p. 5438-5446.
21. Summy, J.M., et al., *The SH4-Unique-SH3-SH2 domains dictate specificity in signaling that differentiate c-Yes from c-Src.* Journal of Cell Science, 2003. **116**(12): p. 2585-2598.
22. Mayer, B.J. and M.J. Eck, *SH3 Domains: Minding your p's and q's.* Current Biology, 1995. **5**(4): p. 364-367.
23. Mayer, B.J., *SH3 domains: complexity in moderation.* Journal of Cell Science, 2001. **114**(7): p. 1253-1263.

24. Waksman, G., et al., *Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed wtih tyrosine-phosphorylated peptides.* Nature, 1992. **358**: p. 646-653.
25. Ladbury, J.E., et al., *Measurement of the binding of tyrosyl phosphopeptides to SH2 domains: a reappraisal.* Proceedings of the National Academy of Sciences, 1995. **92**(8): p. 3199-3203.
26. Taylor, S.S., et al., *Structural Framework for the Protein Kinase Family.* Annual Review of Cell Biology, 1992. **8**(1): p. 429-462.
27. Garcia, P., et al., *Phosphorylation of synthetic peptides containing Tyr-Met-X-Met motifs by nonreceptor tyrosine kinases in vitro.* Journal of Biological Chemistry, 1993. **268**(33): p. 25146-51.
28. Songyang, Z. and L.C. Cantley, *Recognition and specificity in protein tyrosine kinase-mediated signalling.* Trends in Biochemical Sciences, 1995. **20**(11): p. 470-475.
29. Pellicena, P., K.R. Stowell, and W.T. Miller, *Enhanced Phosphorylation of Src Family Kinase Substrates Containing SH2 Domain Binding Sites.* Journal of Biological Chemistry, 1998. **273**(25): p. 15325-15328.
30. Yadav, S.S., et al., *Reengineering the Signaling Properties of a Src Family Kinase.* Biochemistry, 2009. **48**(46): p. 10956-10962.
31. Mayer, B.J. and D. Baltimore, *Mutagenic analysis of the roles of SH2 and SH3 domains in regulation of the Abl tyrosine kinase.* Molecular and Cellular Biology, 1994. **14**(5): p. 2883-2894.
32. Nakamoto, T., et al., *Direct Binding of C-terminal Region of p130 to SH2 and SH3 Domains of Src Kinase.* Journal of Biological Chemistry, 1996. **271**(15): p. 8959-8965.
33. Guappone, A.C., T. Weimer, and D.C. Flynn, *Formation of a stable src–AFAP-110 complex through either an amino-terminal or a carboxy-terminal SH2-binding motif.* Molecular Carcinogenesis, 1998. **22**(2): p. 110-119.
34. Taylor, S.J. and D. Shalloway, *An RNA-binding protein associated with Src through its SH2 and SH3 domains in mitosis.* Nature, 1994. **368**: p. 867-871.
35. Yadav, S.S. and W.T. Miller, *The Evolutionarily Conserved Arrangement of Domains in Src Family Kinases Is Important for Substrate Recognition†.* Biochemistry, 2008. **47**(41): p. 10871-10880.
36. Sicheri, F., I. Moarefi, and J. Kuriyan, *Crystal structure of the Src family tyrosine kinase Hck.* Nature, 1997. **385**(6617): p. 602-609.
37. Xu, W., S.C. Harrison, and M.J. Eck, *Three-dimensional structure of the tyrosine kinase c-Src.* Nature, 1997. **385**: p. 595-602.
38. Yamaguchi, H. and W.A. Hendrickson, *Structural basis for activation of human lymphocyte kinase Lck upon tyrosine phosphorylation.* Nature, 1996. **384**: p. 484-489.
39. Johnson, L.N., M.E.M. Noble, and D.J. Owen, *Active and Inactive Protein Kinases: Structural Basis for Regulation.* Cell, 1996. **85**(2): p. 149-158.
40. Sicheri, F. and J. Kuriyan, *Structures of Src-family tyrosine kinases.* Curr Opin Struct Biol, 1997. **7**(6): p. 777-85.
41. Miller, W.T., *Determinants of Substrate Recognition in Nonreceptor Tyrosine Kinases.* Acc. Chem. Res, 2003. **36**(6): p. 393-400.
42. Moarefi, I., et al., *Activation of the Sire-family tyrosine kinase Hck by SH3 domain displacement.* Nature, 1997. **385**(6617): p. 650-653.
43. LaFevre-Bernt, M., et al., *Intramolecular Regulatory Interactions in the Src Family Kinase Hck Probed by Mutagenesis of a Conserved Tryptophan Residue.* Journal of Biological Chemistry, 1998. **273**(48): p. 32129-32134.
44. Pene-Dumitrescu, T. and T.E. Smithgall, *Expression of a Src Family Kinase in Chronic Myelogenous Leukemia Cells Induces Resistance to Imatinib in a Kinase-dependent Manner.* Journal of Biological Chemistry, 2010. **285**(28): p. 21446-21457.

45. Purushottam, S. and T.E. Smithgall, *Nef alleles from all major HIV-1 clades activate Src-family kinases and enhance HIV-1 replication in an inhibitor-sensitive manner.* PLoS ONE, 2012. **7**(2).
46. Panjarian, S., et al., *Enhanced SH3/Linker Interaction Overcomes Abl Kinase Activation by Gatekeeper and Myristic Acid Binding Pocket Mutations and Increases Sensitivity to Small Molecule Inhibitors.* Journal of Biological Chemistry, 2013. **288**(9): p. 6116-6129.
47. Panjarian, S., et al., *Structure and Dynamic Regulation of Abl Kinases.* Journal of Biological Chemistry, 2013. **288**(8): p. 5443-5450.
48. Osherov, N. and A. Levitzki, *Epidermal-Growth-Factor-Dependent Activation of the Src-Family Kinases.* European Journal of Biochemistry, 1994. **225**(3): p. 1047-1053.
49. Porter, M., et al., *Reciprocal Regulation of Hck Activity by Phosphorylation of Tyr527 and Tyr416 : EFFECT OF INTRODUCING A HIGH AFFINITY INTRAMOLECULAR SH2 LIGAND.* Journal of Biological Chemistry, 2000. **275**(4): p. 2721-2726.
50. Roach, T., et al., *CD45 regulates Src family member kinase activity associated with macrophage integrin-mediated adhesion.* Current biology : CB, 1997. **7**(6): p. 408-417.
51. Okada, M. and H. Nakagawa, *A protein tyrosine kinase involved in regulation of pp60c-src function.* Journal of Biological Chemistry, 1989. **264**(35): p. 20886-93.
52. Michael A Miller, I.A.M., Michael A Shenk, Robert E Steele, *The Src/Csk regulatory circuit arose early in metazoan evolution.* Oncogene, 2000. **19**(34): p. 3925-3930.
53. Suga, H., K. Katoh, and T. Miyata, *Sponge homologs of vertebrate protein tyrosine kinases and frequent domain shufflings in the early evolution of animals before the parazoan–eumetazoan split.* Gene, 2001. **280**(1–2): p. 195-201.
54. Imamoto, A. and P. Soriano, *Disruption of the csk gene, encoding a negative regulator of Src family tyrosine kinases, leads to neural tube defects and embryonic lethality in mice.* Cell, 1993. **73**(6): p. 1117-1124.
55. Nada, S., et al., *Identification of major tyrosine-phosphorylated proteins in Csk-deficient cells.* Oncogene, 1994. **9**(12): p. 3571-8.
56. Thomas, S., P. Soriano, and A. Imamoto, *Specific and redundant roles of Src and Fyn in organizing the cytoskeleton.* Nature, 1995. **376**(6537): p. 267-71.
57. Ogawa, A., et al., *Structure of the Carboxyl-terminal Src Kinase, Csk.* Journal of Biological Chemistry, 2002. **277**(17): p. 14351-14354.
58. Lamers, M.B.A.C., et al., *Structure of the protein tyrosine kinase domain of C-terminal Src kinase (CSK) in complex with staurosporine.* Journal of Molecular Biology, 1999. **285**(2): p. 713-725.
59. Sondhi, D. and P.A. Cole, *Domain Interactions in Protein Tyrosine Kinase Csk†.* Biochemistry, 1999. **38**(34): p. 11147-11155.
60. Shekhtman, A., et al., *Novel mechanism of regulation of the non-receptor protein tyrosine kinase csk: insights from NMR mapping studies and site-directed mutagenesis.* Journal of Molecular Biology, 2001. **314**(1): p. 129-138.
61. Chong, Y.-P., T.D. Mulhern, and H.-C. Cheng, *C-terminal Src kinase (CSK) and CSK-homologous kinase (CHK)—endogenous negative regulators of Src-family protein kinases.* Growth Factors, 2005. **23**(3): p. 233-244.
62. Kawabuchi, M., et al., *Transmembrane phosphoprotein Cbp regulates the activities of Src-family tyrosine kinases.* Nature, 2000. **404**(6781): p. 999-1003.
63. Cao, H., W.E. Courchesne, and C.C. Mastick, *A Phosphotyrosine-dependent Protein Interaction Screen Reveals a Role for Phosphorylation of Caveolin-1 on Tyrosine 14: RECRUITMENT OF C-TERMINAL Src KINASE.* Journal of Biological Chemistry, 2002. **277**(11): p. 8771-8774.
64. Tsutsumi, R., et al., *Attenuation of Helicobacter pylori CagA·SHP-2 Signaling by Interaction between CagA and C-terminal Src Kinase.* Journal of Biological Chemistry, 2003. **278**(6): p. 3664-3670.

65. Levinson, N.M., P. Visperas, and J. Kuriyan, *The tyrosine kinase Csk dimerizes through its SH3 domain.* PLoS ONE, 2009. **4**(11).
66. Bergman, M., et al., *The human p50csk tyrosine kinase phosphorylates p56Ick at Tyr-505 and down regulates its catalytic activity.* EMBO Journal, 1992. **11**(8): p. 2919-24.
67. Sondhi, D., et al., *Peptide and Protein Phosphorylation by Protein Tyrosine Kinase Csk: Insights into Specificity and Mechanism†.* Biochemistry, 1998. **37**(1): p. 165-172.
68. Chong, Y.-P., et al., *A Novel Non-catalytic Mechanism Employed by the C-terminal Src-homologous Kinase to Inhibit Src-family Kinase Activity.* Journal of Biological Chemistry, 2004. **279**(20): p. 20752-20766.
69. Levinson, N.M., et al., *Structural Basis for the Recognition of c-Src by Its Inactivator Csk.* Cell, 2008. **134**(1): p. 124-134.
70. Manning, G., et al., *The Protein Kinase Complement of the Human Genome.* Science, 2002. **298**(5600): p. 1912-1934.
71. Xu, B.-e., et al., *WNK1, a Novel Mammalian Serine/Threonine Protein Kinase Lacking the Catalytic Lysine in Subdomain II.* Journal of Biological Chemistry, 2000. **275**(22): p. 16795-16801.
72. Min, X., et al., *Crystal Structure of the Kinase Domain of WNK1, a Kinase that Causes a Hereditary Form of Hypertension.* Structure, 2004. **12**(7): p. 1303-1311.
73. Mukherjee, K., et al., *CASK Functions as a Mg2+-Independent Neurexin Kinase.* Cell, 2008. **133**(2): p. 328-339.
74. Mukherjee, K., et al., *Evolution of CASK into a Mg2+-Sensitive Kinase.* Sci. Signal., 2010. **3**(119): p. ra33-.
75. Zhang, H., et al., *The role of pseudokinases in cancer.* Cellular Signalling, 2012. **24**(6): p. 1173-1184.
76. Ruiz-Trillo, I., et al., *A Phylogenomic Investigation into the Origin of Metazoa.* Molecular Biology and Evolution, 2008. **25**(4): p. 664-672.
77. King, N., *The unicellular ancestry of animal development.* Dev. Cell, 2004. **7**: p. 313-325.
78. Rokas, A., *The molecular origins of multicellular transitions.* Curr Opin Genet Dev, 2008. **18**(6): p. 472-8.
79. Lim, W.A. and T. Pawson, *Phosphotyrosine signaling: evolving a new cellular communication system.* Cell, 2010. **142**(5): p. 661-7.
80. King, N., et al., *The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans.* Nature, 2008. **451**(7180): p. 783-788.
81. Pincus, D., et al., *Evolution of the phospho-tyrosine signaling machinery in premetazoan lineages.* Proceedings of the National Academy of Sciences, 2008. **105**(28): p. 9680-9684.
82. Manning, G., et al., *The protist, Monosiga brevicollis, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan.* Proceedings of the National Academy of Sciences, 2008. **105**(28): p. 9674-9679.
83. Miller, W.T., *Tyrosine kinase signaling and the emergence of multicellularity.* Biochimica et Biophysica Acta (BBA) - Molecular Cell Research, 2012. **1823**(6): p. 1053-1057.
84. Leadbeater, B.S.C., *Life-history and ultrastructure of a new marine species of Proterospongia (Choanoflagellida).* Journal of the Marine Biological Association of the United Kingdom, 1983. **63**(01): p. 135-160.
85. King, N. and S.B. Carroll, *A receptor tyrosine kinase from choanoflagellates: Molecular insights into early animal evolution.* Proceedings of the National Academy of Sciences, 2001. **98**(26): p. 15032-15037.
86. King, N., C.T. Hittinger, and S.B. Carroll, *Evolution of key cell signaling and adhesion protein families predates animal origins.* Science, 2003. **301**: p. 361-363.

87. Segawa, Y., et al., *Functional development of Src tyrosine kinases during evolution from a unicellular ancestor to multicellular animals.* Proceedings of the National Academy of Sciences, 2006. **103**(32): p. 12021-12026.
88. Snell, E.A., et al., *An unusual choanoflagellate protein released by Hedgehog autocatalytic processing.* Proceedings of the Royal Society B: Biological Sciences, 2006. **273**(1585): p. 401-407.
89. Lavrov, D.V., et al., *Mitochondrial Genomes of Two Demosponges Provide Insights into An Early Stage of Animal Evolution.* Molecular Biology and Evolution, 2005. **22**(5): p. 1231-1239.
90. Rokas, A., D. Krüger, and S.B. Carroll, *Animal Evolution and the Molecular Signature of Radiations Compressed in Time.* Science, 2005. **310**(5756): p. 1933-1938.
91. Dayel, M.J., et al., *Cell differentiation and morphogenesis in the colony-forming choanoflagellate Salpingoeca rosetta.* Developmental Biology, 2011. **357**(1): p. 73-82.
92. Fairclough, S.R., M.J. Dayel, and N. King, *Multicellular development in a choanoflagellate.* Current Biology, 2010. **20**(20): p. R875-R876.
93. Li, W., et al., *Signaling Properties of a Non-metazoan Src Kinase and the Evolutionary History of Src Negative Regulation.* Journal of Biological Chemistry, 2008. **283**(22): p. 15491-15501.
94. Li, W., S. Scarlata, and W.T. Miller, *Evidence for Convergent Evolution in the Signaling Properties of a Choanoflagellate Tyrosine Kinase.* Biochemistry, 2009. **48**(23): p. 5180-5186.
95. Ruiz-Trillo, I., et al., *Capsaspora owczarzaki is an independent opisthokont lineage.* Current Biology, 2004. **14**(22): p. R946-R947.
96. Ruiz-Trillo, I., et al., *Insights into the evolutionary origin and genome architecture of the unicellular opisthokonts Capsaspora owczarzaki and Sphaeroforma arctica.* J. Eukaryot. Microbiol., 2006. **53**: p. 379-384.
97. Shalchian-Tabrizi, K., et al., *Multigene Phylogeny of Choanozoa and the Origin of Animals.* PLoS ONE, 2008. **3**(5): p. e2098.
98. Suga, H., et al., *Genomic survey of premetazoans shows deep conservation of cytoplasmic tyrosine kinases and multiple radiations of receptor tyrosine kinases.* Sci Signal, 2012. **5**(222): p. ra35.
99. Owczarzak, A., H.H. Stibbs, and C.J. Bayne, *The destruction of Schistosoma mansoni mother sporocysts in vitro by amoebae isolated from Biomphalaria glabrata: an ultrastructural study.* Journal of Invertebrate Pathology, 1980. **35**(1): p. 26-33.
100. Casnellie, J.E., *Assay of protein kinases using peptides with basic residues for phosphocellulose binding.* Methods Enzymol, 1991. **200**: p. 115-20.
101. Bonner, J., *The Origins of Multicellularity.* Integrative Biology, 1998. **1**: p. 27-36.
102. Fantl, W.J., D.E. Johnson, and L.T. Williams, *Signalling by Receptor Tyrosine Kinases.* Annual Review of Biochemistry, 1993. **62**(1): p. 453-481.
103. Gerhart, J., *1998 warkany lecture: Signaling pathways in development.* Teratology, 1999. **60**(4): p. 226-239.
104. Hertel, L.A., C.J. Bayne, and E.S. Loker, *The symbiont Capsaspora owczarzaki, nov. gen. nov. sp., isolated from three strains of the pulmonate snail Biomphalaria glabrata is related to members of the Mesomycetozoea.* International Journal for Parasitology, 2002. **32**(9): p. 1183-1191.
105. Sebé-Pedrós, A., et al., *Ancient origin of the integrin-mediated adhesion and signaling machinery.* Proceedings of the National Academy of Sciences, 2010. **107**(22): p. 10142-10147.
106. Scott, M.P. and W.T. Miller, *A Peptide Model System for Processive Phosphorylation by Src Family Kinases†.* Biochemistry, 2000. **39**(47): p. 14531-14537.
107. Klinghoffer, R., et al., *Src family kinases are required for integrin but not PDGF signal transduction.* EMBO Journal, 1999. **18**: p. 2459-2471

108. Lee, S., et al., *Determination of the substrate-docking site of protein tyrosine kinase C-terminal Src kinase.* Proc Natl Acad Sci U S A, 2003. **100**(25): p. 14707-12.
109. Mikkola, E.T. and C.G. Gahmberg, *Hydrophobic Interaction between the SH2 Domain and the Kinase Domain Is Required for the Activation of Csk.* Journal of Molecular Biology, 2010. **399**(4): p. 618-627.
110. Lin, X., et al., *Structural Basis for Domain–Domain Communication in a Protein Tyrosine Kinase, the C-terminal Src Kinase.* Journal of Molecular Biology, 2006. **357**(4): p. 1263-1273.
111. Hanks, S.K. and T. Hunter, *Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification.* The FASEB Journal, 1995. **9**(8): p. 576-96.
112. Khersonsky, O., C. Roodveldt, and D.S. Tawfik, *Enzyme promiscuity: evolutionary and mechanistic aspects.* Current Opinion in Chemical Biology, 2006. **10**(5): p. 498-508.
113. Scheeff, E.D., et al., *Structure of the Pseudokinase VRK3 Reveals a Degraded Catalytic Site, a Highly Conserved Kinase Fold, and a Putative Regulatory Binding Site.* Structure (London, England : 1993), 2009. **17**(1): p. 128-138.
114. Boudeau, J., et al., *Emerging roles of pseudokinases.* Trends in Cell Biology, 2006. **16**(9): p. 443-452.
115. Kruse, M., I.M. Müller, and W.E. Müller, *Early evolution of metazoan serine/threonine and tyrosine kinases: identification of selected kinases in marine sponges.* Molecular Biology and Evolution, 1997. **14**(12): p. 1326-1334.
116. Lang, B.F., et al., *The Closest Unicellular Relatives of Animals.* Current Biology, 2002. **12**(20): p. 1773-1778.
117. Steenkamp, E.T., J. Wright, and S.L. Baldauf, *The Protistan Origins of Animals and Fungi.* Molecular Biology and Evolution, 2006. **23**(1): p. 93-106.
118. Fairclough, S., et al., *Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate Salpingoeca rosetta.* Genome Biology, 2013. **14**(2): p. R15.
119. Horvath, C.M., Z. Wen, and J.E. Darnell, *A STAT protein domain that determines DNA sequence recognition suggests a novel DNA-binding domain.* Genes & Development, 1995. **9**(8): p. 984-994.
120. Darnell, J.E., *STATs and Gene Regulation.* Science, 1997. **277**(5332): p. 1630-1635.
121. Heim, M., et al., *Contribution of STAT SH2 groups to specific interferon signaling by the Jak-STAT pathway.* Science, 1995. **267**(5202): p. 1347-1349.
122. Chen, X., et al., *Crystal Structure of a Tyrosine Phosphorylated STAT-1 Dimer Bound to DNA.* Cell, 1998. **93**(5): p. 827-839.
123. Callebaut, I., et al., *HYR, an extracellular module involved in cellular adhesion and related to the immunoglobulin-like fold.* Protein Science, 2000. **9**(7): p. 1382-1390.
124. Favelyukis, S., et al., *Structure and autoregulation of the insulin-like growth factor I receptor kinase.* Nature Structural and Molecular Biology, 2001. **8**: p. 1058-1063.
125. Z, S., et al., *Catalytic specificity of protein-tyrosine kinases is critical for selective signalling.* Nature, 1993. **373**: p. 536-539.
126. Suga, H., et al., *Intermittent divergence of the protein tyrosine kinase family during animal evolution.* FEBS Letters, 1997. **412**(3): p. 540-546.
127. Jin, J. and T. Pawson, *Modular evolution of phosphorylation-based signalling systems.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2012. **367**(1602): p. 2540-2555.
128. Srivastava, M., et al., *The Amphimedon queenslandica genome and the evolution of animal complexity.* Nature, 2010. **466**(7307): p. 720-726.
129. Nichols, S.A., et al., *Early evolution of animal cell signaling and adhesion genes.* Proceedings of the National Academy of Sciences, 2006. **103**(33): p. 12451-12456.

130. Taylor, S.S., et al., *Evolution of the eukaryotic protein kinases as dynamic molecular switches.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2012. **367**(1602): p. 2517-2528.
131. Schultheiss, K.P., et al., *Metazoan-like signaling in a unicellular receptor tyrosine kinase.* BMC Biochemistry, 2013. **14**(4).
132. Hörtner, M., et al., *A new high affinity binding site for suppressor of cytokine signaling-3 on the erythropoietin receptor.* European Journal of Biochemistry, 2002. **269**(10): p. 2516-2526.
133. Hitchcock, I.S., et al., *YRRL motifs in the cytoplasmic domain of the thrombopoietin receptor regulate receptor internalization and degradation.* Blood, 2008. **112**(6): p. 2222-2231.
134. Gelderloos, J.A., et al., *A Role for Src in Signal Relay by the Platelet-derived Growth Factor α Receptor.* Journal of Biological Chemistry, 1998. **273**(10): p. 5908-5915.
135. Kypta, R.M., et al., *Association between the PDGF receptor and members of the src family of tyrosine kinases.* Cell, 1990. **62**(3): p. 481-492.
136. Horvath, C.M., *STAT proteins and transcriptional responses to extracellular signals.* Trends in Biochemical Sciences, 2000. **25**(10): p. 496-502.
137. Benekli, M., et al., *Signal transducer and activator of transcription proteins in leukemias.* Blood, 2003. **101**(8): p. 2940-2954.
138. Reich, N. and L. Liu, *Tracking STAT nuclear traffic.* Nature Reviews Immunology, 2006. **6**: p. 602-612.
139. Weinmaster, G. and T. Pawson, *Protein kinase activity of FSV (Fujinami sarcoma virus) P130gag-fps shows a strict specificity for tyrosine residues.* Journal of Biological Chemistry, 1986. **261**(1): p. 328-33.
140. Taylor, S.S., E. Radzio-Andzelm, and T. Hunter, *How do protein kinases discriminate between serine/threonine and tyrosine? Structural insights from the insulin receptor protein-tyrosine kinase.* The FASEB Journal, 1995. **9**(13): p. 1255-66.
141. Torruella, G., et al., *Phylogenetic Relationships within the Opisthokonta Based on Phylogenomic Analyses of Conserved Single-Copy Protein Domains.* Molecular Biology and Evolution, 2012. **29**(2): p. 531-544.
142. King, N., *The Unicellular Ancestry of Animal Development.* Developmental Cell, 2004. **7**(3): p. 313-325.
143. Akst, J., *From Simple to Complex.* Science, 2011.