5-6-2017

Toward the Characterization of Mammalian Sterile20-like Kinase 3 (MST3) and the Conserved Herpesviridae Protein Kinases (CHPKs) through Phosphorylation Motifs

Amanda S. Wildstein
University of Connecticut - Storrs, amanda.wildstein@uconn.edu

Toward the Characterization of Mammalian Sterile20-like Kinase 3 (MST3) and the Conserved Herpesviridae Protein Kinases (CHPKs) through Phosphorylation Motifs

Amanda Susan Wildstein

B.S., University of Connecticut, 2016

A Thesis

Submitted in Partial Fulfillment of the

Requirements of the Degree of

Master of Science

at the

University of Connecticut

2017

APPROVAL PAGE

Master of Science Thesis

Toward the Characterization of Mammalian Sterile20-like Kinase 3 (MST3) and the Conserved Herpesviridae Protein Kinases (CHPKs) through Phosphorylation Motifs

Presented by

Amanda Susan Wildstein, B.S.

Major Advisor _____
Daniel Schwartz

Associate Advisor _____
Joseph LoTurco

Associate Advisor _____
John Redden

University of Connecticut

2017

ACKNOWLEDGEMENTS

First and foremost, I want to thank my advisor Dr. Daniel Schwartz. I appreciate all of his contributions of time, guidance and funding, without which this work would not have been possible. He always allowed my research to be my own while providing direction whenever I needed it. I would also to thank Drs. Joseph LoTurco and John Redden for serving on my committee.

Additionally, I would like to acknowledge members of the Schwartz lab, both past and present, for building the computational tools, optimizing protocols and collaborating with me. I could not have completed this without their contributions. I would like to especially thank Joshua Lubner for his endless patience and continued support. He has helped me at every step along the way from teaching me how to pipette to editing the final drafts of this thesis.

Finally, I want to express my gratitude to my closest friends and family for their continued support and encouragement.

TABLE OF CONTENTS

Introduction

Protein kinases represent one of the largest superfamilies of proteins with more than 500 members. Collectively, they account for about 2% of all identified human genes[1]. The substantial number of kinases present in humans may be explained by the fact that they mediate a majority of signal transduction in cells and are integral regulators in a myriad of pathways including metabolism, growth, and replication. Kinases influence these (and other) pathways by post-translationally modifying their substrates. More specifically, kinases catalyze the addition of the gamma phosphate group from ATP to a phosphoacceptor amino acid (most commonly a serine, threonine, or tyrosine residue in eukaryotic cells)[2]. The addition of a negatively charged phosphate group to a protein can have a significant effect on its enzymatic activity, subcellular localization, local or global conformation, and/or interacting partners[3]. This range of effects that phosphorylation may exert on a substrate has made kinases key components of many biochemical processes.

Since phosphorylation can affect a protein and thus cellular processes so dramatically, it is of critical importance that a kinase be specific in its activity. A kinase must distinguish between its substrate and the countless other proteins it may encounter. Additionally, once a kinase finds its substrate it has to further discern which specific residue to phosphorylate. The precision in phosphorylation that is required for normal cellular functioning is much too great to leave kinase-substrate interactions to chance; in fact, aberrant kinase activity is the cause of many human diseases and cancers[4,5]. As such there are multiple ways in which specificity is conferred. On the most basic level, spatial and temporal regulation prevents incorrect physical interactions between kinase and substrate[6,7]. Once the kinase and substrate are expressed in the same cellular compartment of the same cell, and at the same time, potential phosphorylation substrates are further narrowed down by their ability to form energetically favorable connections

with the kinase's catalytic domain. Oftentimes, the connections occur between complementary amino acids of the substrate and kinase on the basis of hydrogen bonds, ionic attractions, and hydrophobic effects[8]. In this way, the primary structure of the substrate surrounding one of its phosphoacceptor residues creates a specific biochemical environment that can be recognized by the active site of a particular kinase and thus catalyze phosphorylation of that site.

More closely examining the primary structure of a kinase's substrates can yield valuable information for determining its specificity. We can organize the phosphosites by aligning them with the phosphoacceptor as the central residue to facilitate the identification of patterns in primary structure. A pattern like this can be described as a motif, which is, fundamentally, a representation of the biochemical environment that helps to direct the kinase to the phosphoacceptor. Motifs are powerful tools for predicting site-specific interactions between kinase and substrate, and can be an important first step in deciphering complicated biochemical pathways[9]. Motifs can be extracted by eye by aligning substrates, specifically the residues around the phosphoacceptor. For example, aligning some of the substrates of PKA reveals a clear preference for substrates with arginine 3 positions upstream of the phosphoacceptor (the so-called P-3 position) and another positively charged amino acid following it (at P-2, **Figure 1**).

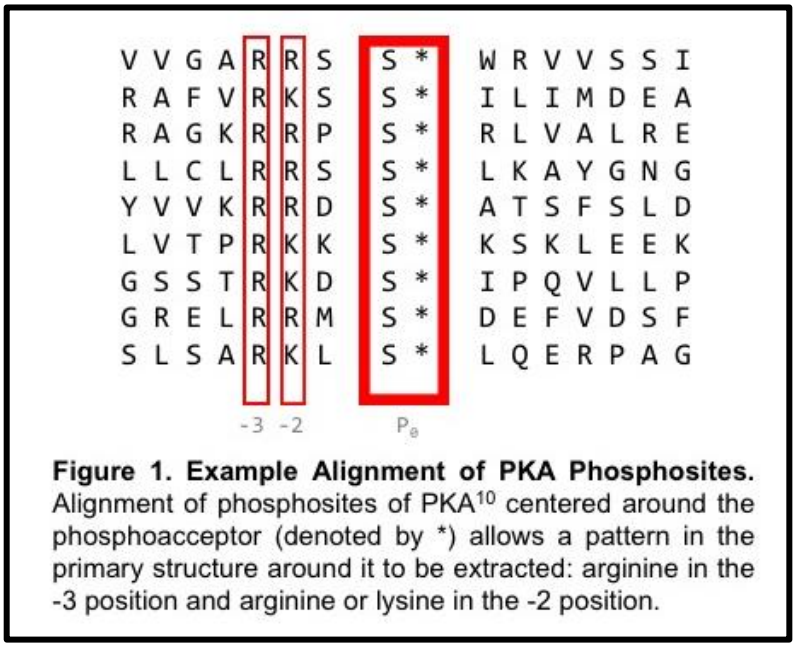

Figure 1. Example Alignment of PKA Phosphosites. Alignment of phosphosites of PKA[10] centered around the phosphoacceptor (denoted by *) allows a pattern in the primary structure around it to be extracted: arginine in the -3 position and arginine or lysine in the -2 position.

Unlike the example in **Figure 1**, it is often impossible to derive accurate motifs from the alignment of a few 15-mers as not all substrates exhibit the motif. It is equally unfeasible to identify patterns by eye in a large data set of hundreds of phosphopeptides. Generating a statistically significant and accurate motif may require the alignment of hundreds of unique substrates; but there is a very limited knowledge of physiologic substrates for the majority of protein kinases due to the lack of an efficient means to discover them[10] (**Figure 2**). It is equally difficult to pair substrates or known phosphorylated residues to their direct upstream kinase as signaling cascades are often very complex and involve many kinases. Thus, the traditional method of determining motifs does not rely on identifying proteins phosphorylated by a given kinase but rather relies upon identifying peptides that a kinase will phosphorylate *in vitro*. These methods most commonly involve incubating recombinant kinase and radiolabeled ATP with a combinatorial peptide library *in vitro*[11]. The peptides in the library are arranged in a grid and each well contains peptides that have a fixed phosphoacceptor and one other fixed residue. All other positions are degenerate in order to evaluate the influence of the one fixed residue. The peptides are transferred to a membrane and radioactivity (from phosphorylation by the kinase of interest with radiolabeled ATP) is measured quantitatively. The motif can be read by looking at the largest incorporation of radioactive phosphate at a given position. This approach is limited by the need for recombinant kinase, the use of a radioactive isotope, the cost of a sufficiently large synthetic peptide library, the use of peptides rather than more physiologically relevant proteins, and a loss of correlation between positions. In order to overcome some of the limitations of the peptide library approach, we previously developed a novel *in vivo* proteomic methodology to quickly, accurately and cost effectively determine kinase specificity[12].

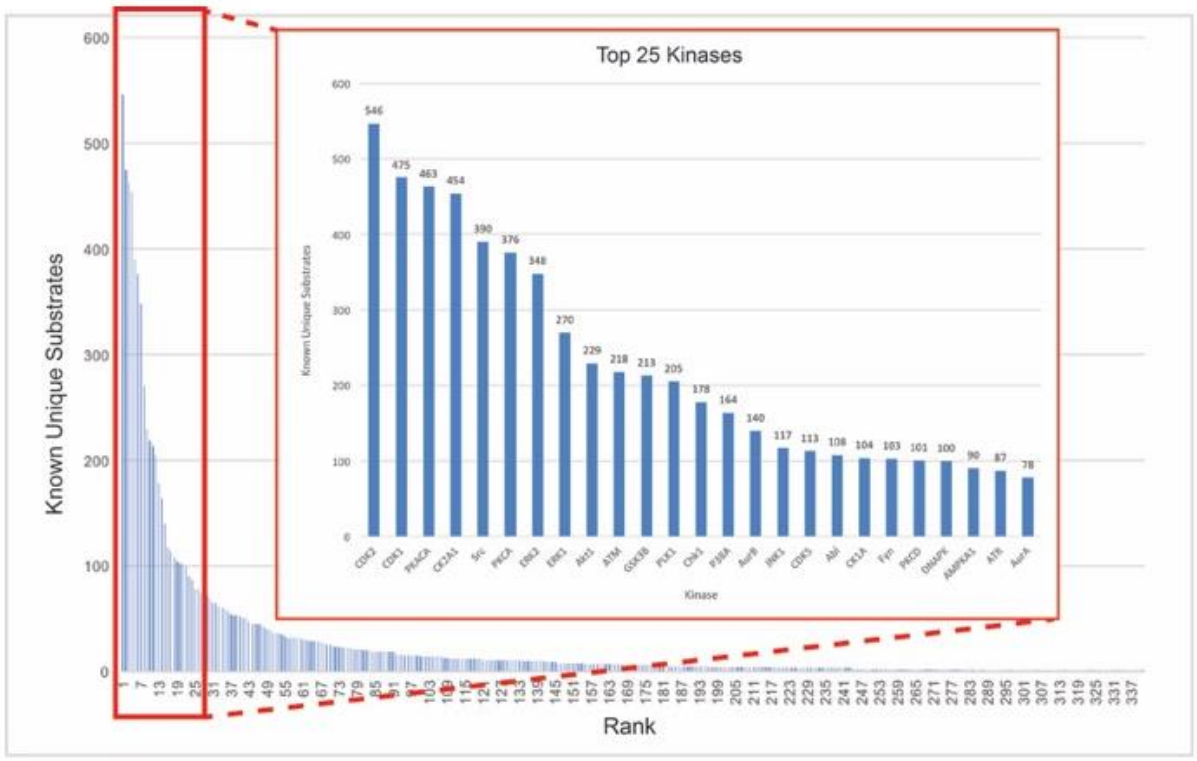

Figure 2. Number of Known Substrates per Kinase. Data from PhosphoSitePlus[10] shows the number of unique substrates identified per kinase drops precipitously after the top 10 most well characterized human kinases.

ProPeL

Our lab has developed a novel, *in vivo* methodology to interrogate kinase specificity called ProPeL[12], for Proteomic Peptide Library (**Figure 3**). ProPeL combines the ease of using bacterial transcription and translation machinery to produce the kinase of interest (to avoid the need to purify large amounts of recombinant kinase) with the analytical power of liquid chromatography tandem mass spectrometry to identify phosphopeptides. This approach also makes use of the bacterial proteome, which is sufficiently large and diverse to be used in the same way as a more traditional combinatorial peptide library. The endogenous bacterial proteins are phosphorylated by the exogenous kinase of interest according to its native specificity. *E. coli* is an appropriate host organism for this method because it typically has very low levels of endogenous serine/threonine/tyrosine phosphorylation[12–14] meaning that one can

assume that the phosphorylated residues are the direct result of the exogenous kinase and not an *E. coli* kinase. This is absolutely crucial to the methodology and gives us a direct link between kinase and substrate due to the absence of confounding serine/threonine/tyrosine kinases. There is also very high signal to noise ratio in our data due to the low levels of background phosphorylation in *E. coli*. Additionally, the ProPeL approach allows for phosphorylation to occur with full-length protein substrates (rather than peptides), and under more physiologically relevant conditions than an *in vitro* assay. After expression of the kinase of interest in *E. coli*, phosphopeptides can be identified by LC-MS/MS. These sites are then analyzed computationally using the Probability Logo (pLogo) Generator created by our lab[15] to visualize specificity motif(s) (see **Materials and Methods** for details). Here, we apply that methodology to discover the specificity motif of Mammalian Sterile 20-Like Kinase 3 (MST3) and the 8 Conserved Herpesvirus Protein Kinases.

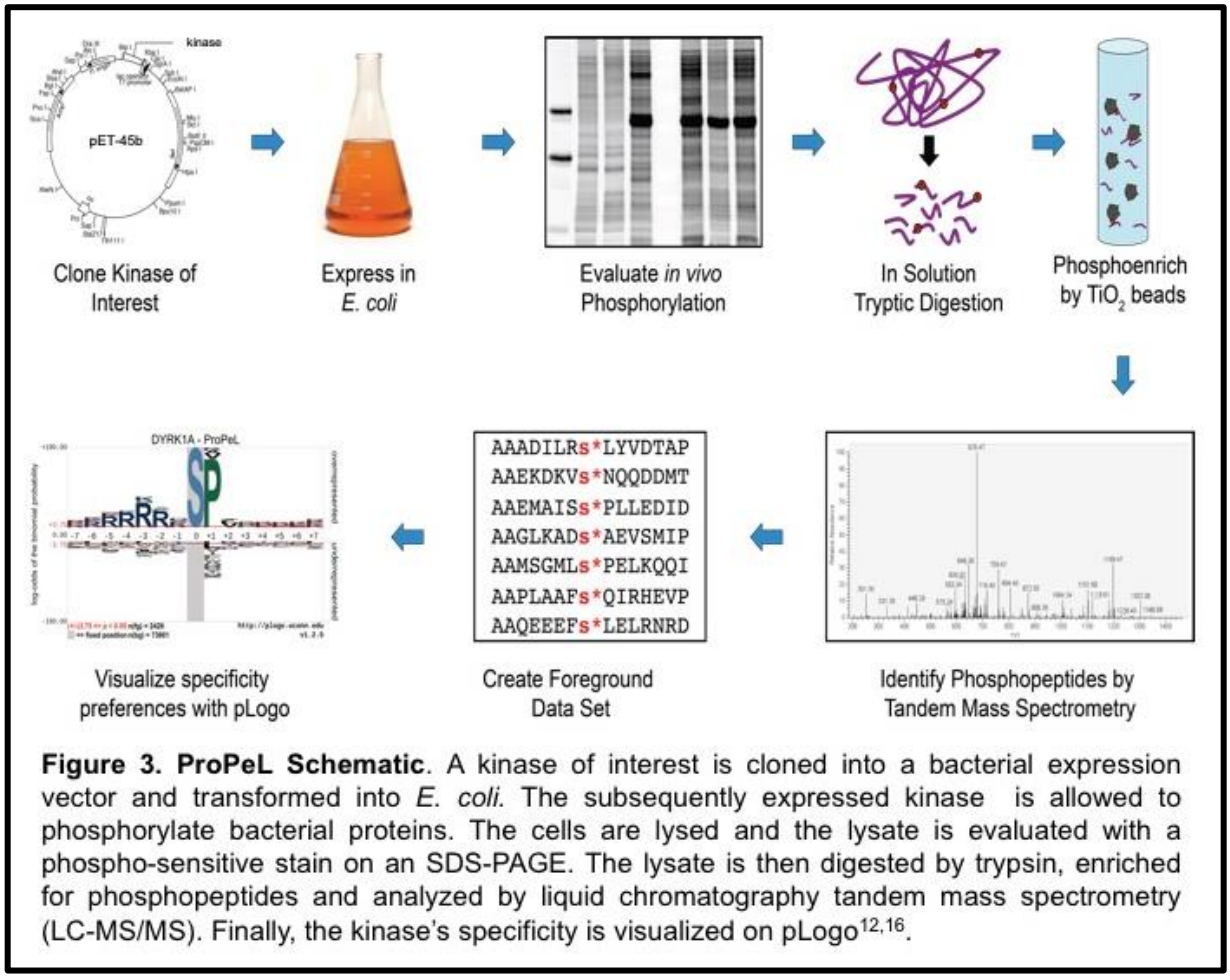

Figure 3. ProPeL Schematic. A kinase of interest is cloned into a bacterial expression vector and transformed into *E. coli*. The subsequently expressed kinase is allowed to phosphorylate bacterial proteins. The cells are lysed and the lysate is evaluated with a phospho-sensitive stain on an SDS-PAGE. The lysate is then digested by trypsin, enriched for phosphopeptides and analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS). Finally, the kinase's specificity is visualized on pLogo[12,16].

Project One: Discovering the Specificity Motif of Mammalian Sterile 20-like Kinase 3 (MST3)

The Mammalian Sterile 20-like (MST) family of human kinases is made up of 5 kinases. They are all very closely related to the well-studied Hippo kinase in *D. melanogaster* and, like Hippo, have been broadly implicated in the regulation of cell proliferation, organ size, cell migration and cell polarity[16]. Our research into this family is focused on Mammalian Sterile 20-like Kinase 3 (MST3), a ubiquitously expressed serine/threonine protein kinase activated primarily by caspase cleavage. Recent studies have illustrated the importance of MST3 in pathways regulating cytoskeletal elements of neurons. MST3 has been shown to be necessary for the maintenance of dendritic spines in excitatory neurons[17], axonal response to nerve growth factor[18] and migration of pyramidal neurons to the cortical plate during development[19]. The loss of function of MST3 has also been shown to reduce activity at glutamatergic synapses[19]. Studies suggest disruption of excitatory pathways may be involved in the development of intellectual disabilities such as Autism Spectrum Disorders[20]. It is our hope that characterizing the specificity of MST3 and predicting its potential substrates will shed light on these pathways.

Published data, based on a very limited number of phosphosites obtained from an *in vitro* ATP cross-linking assay, suggest that the specify motif of MST3 is T*ΦR (where * denotes the phosphoacceptor, and Φ denotes any hydrophobic residue)[17]. There are no serine sites currently reported in the literature, though it is likely that MST3 phosphorylates both serine and threonine *in vivo* as other members of the Sterile 20-like family do. It is important to note that the sites used to generate the specificity motif in the pLogo in **Figure 4** were obtained using a mutant MST3 whose ATP-binding pocket was modified to accommodate an ATP-γ-S analog[21] which, due to its critical role in catalysis, may have affected specificity. Thus, we applied the

ProPeL methodology to MST3 to obtain more phosphosites and thus a more robust and accurate motif.

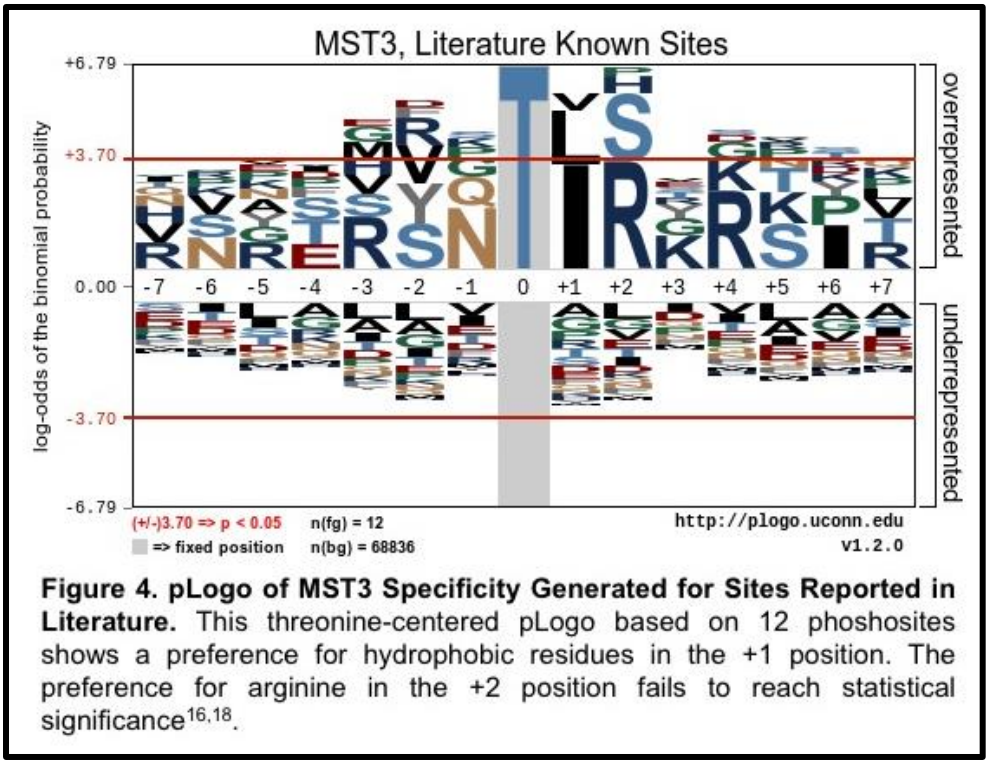

Figure 4. pLogo of MST3 Specificity Generated for Sites Reported in Literature. This threonine-centered pLogo based on 12 phoshosites shows a preference for hydrophobic residues in the +1 position. The preference for arginine in the +2 position fails to reach statistical significance[16,18].

Project Two: Discovering the Specificity Motif of the Conserved Herpesviridae Protein Kinases

Together the 8 human Herpesviridae are responsible for millions of new infections each year. Symptoms of Herpes viral infections can range from relatively benign cutaneous lesions to life-threatening encephalitis. After an acute initial infection, all herpes viruses establish latency and can reactivate at anytime and cause symptoms of varying severity. There is currently no cure for any Herpesvirus infection, only treatments to shorten the duration of symptoms. A promising therapeutic target may be the Conserved Herpesviridae Protein Kinases (CHPKs), orthologs of the UL13 kinase from HHV-1/2, of which each virus expresses an isoform (see **Table 1**).

CHPKs are essential to viral replication, nuclear egress[22] and the establishment of latency[23]. Evidence suggests that they may behave similarly to cyclin dependent kinases[23–25]; however, there is very little known about the specific pathways in which they participate once expressed in the host's cells. A clear motif for these proteins will serve as a starting point to predict cellular substrates and thus provide insight into their physiologic roles. CHPKs represent promising drug targets due to their important role in virulence and low homology with human kinases, which should limit off-target effects of potential CHPK inhibitors. It is our ultimate goal that a better understanding of these viral kinases will help to direct targeted therapeutic development[26].

Virus	CHPK Ortholog
Herpes Simplex Virus 1 (HSV-1/HHV-1)	UL13
Herpes Simplex Virus 2 (HSV-2/HHV-2)	UL13
Varicella Zoster Virus (VZV/HHV-3)	Orf47
Epstein-Barr Virus (EBV/HHV-4)	BGLF4
Human Cytomegalovirus (HCMV/HHV-5)	UL97
Roseolovirus (HHV-6)	U69
Human Herpes Virus 7 (HHV-7)	U69
Kaposi's Sarcoma-Associated Herpes Virus (KSHV/HHV-8)	Orf36

Table 1. Conserved Herpesviridae Protein Kinase Orthologs. All human herpesviruses encode a CHPK ortholog[24].

Here we sought to apply ProPeL to the entire CHPK family, representing the first application of the methodology to non-eukaryotic kinases. The regions outside of the kinase domains of all 8 CHPKs are poorly characterized[27] leaving the possibility open for other mechanisms that impart specificity including non-catalytic functions such as mediating protein-protein interactions. It is also worth noting that the activation segment, a region important for catalytic activity and confer

substrate specificity motifs[28], is poorly conserved between CHPKs as evidenced by a multiple sequence alignment (**Figure 9**). The DFG and APE motifs, which mark the activation segment of a kinase, are highly conserved among eukaryotic protein kinases[29], but are poorly conserved within the CHPK family. Notably, the critical magnesium-chelating aspartate of the canonical DFG motif is conserved across all CHPKs and the phenylalanine of the DFG motif is maintained or replaced by another aromatic amino acid. However, the APE domain, which nucleates substrate binding, is almost nonexistent in the CHPKs. It is unknown what effect this may have

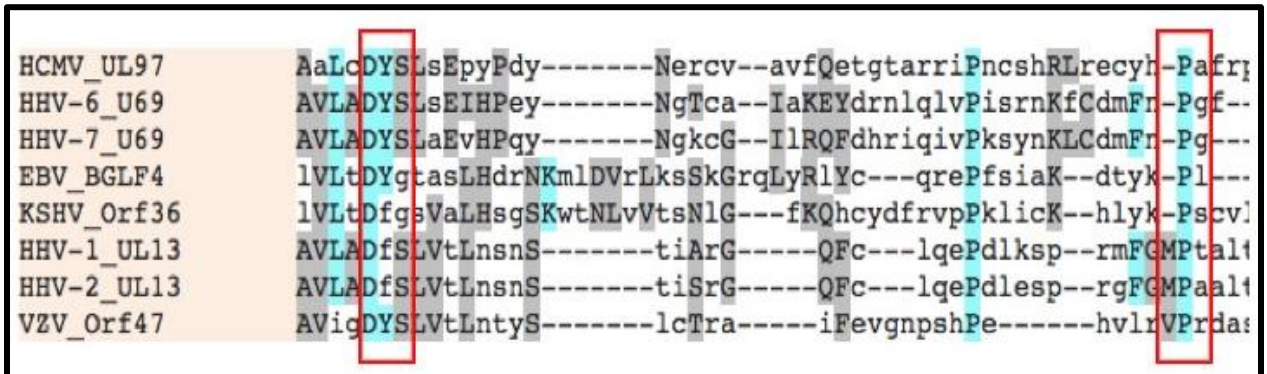

Figure 5. CHPK Activation Segment Alignment. The DFG and APE motifs mark the activation segment of a kinase, an important region for catalytic activity. They are highly conserved between eukaryotic kinases[30], but are poorly conserved between CHPKs as marked by the red boxes.

on the activity and specificity of these kinases. We hope that our ProPeL approach will begin to shed light on this.

Results

MST3

Optimizing MST3 Expression

In order to use ProPeL to determine the specificity motif of MST3, a constitutively active truncation of MST3 was expressed in *E. coli*. The truncation contains the first 325 amino acids and lacks its auto-inhibitory C-terminus which mimics the activation by caspases that occurs normally[30]. The MST3 construct was cloned into the bacterial expression vector pET45b using

overlap extension[31]. The construct was then transformed into C41(DE3) cells which exhibit increased ability to express exogenous, toxic proteins[32].

Expression of the kinase in *E. coli* was a significant point of optimization to maximize soluble, catalytically active kinase yield. After every expression condition the cells were lysed and the expression, auto-phosphorylation and total activity of the kinase were evaluated using SDS-PAGE with the in-gel phosphoprotein Pro-Q Diamond stain, and normalized to the total amount of protein shown with Coomassie staining. If a kinase is catalytically active, there will be more phosphorylation across a range of molecular weights and a band at the molecular weight of the kinase showing that it has been autophosphorylated (a requirement for enzymatic activity).

There are many variables one can manipulate to increase heterologous protein expression. The temperature of a 16 hour incubation was varied as reducing the incubation temperature below 37ºC can increase yields by slowing the rate of growth of the *E. coli* and overall metabolic stress as well as promote proper folding of the kinase[33]. These samples were induced at the mid-log stage of growth, evaluated by optical density (OD_{600} = 0.4-0.6), with 0.5 mM IPTG (all samples were induced this way unless otherwise noted) and incubated at either 20°C or 37°C. Neither of these conditions showed significant expression (**Figure 6**). The point of induction was also changed to the very start of the incubation without success (**Figure 7, lanes 1-2**). Next, the incubation duration was varied because shortening the incubation may be helpful if the kinase of interest is particularly cytotoxic; whereas a longer incubation at a lower temperature may increase yields of a large kinase that is expressed at low levels due to the metabolic demands its translation places on the cell. These changes however, did not yield detectable amount of MST3 either (**Figure 7, lanes 3-6**).

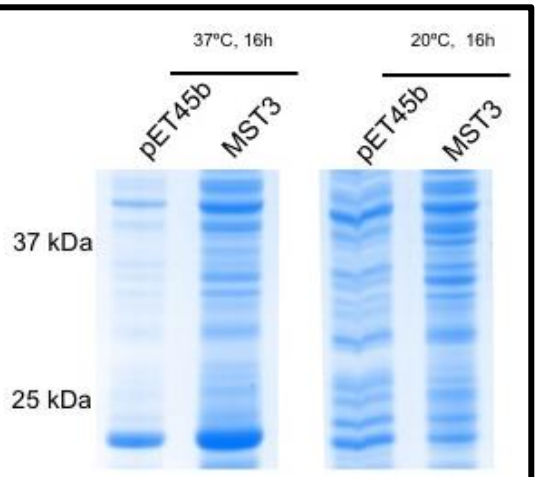

Figure 6. Optimization of MST3 Expression Conditions. We evaluated expression of MST3 (36kDa) in multiple growth conditions using SDS-PAGE. Coomassie staining shows no significant expression of MST3.

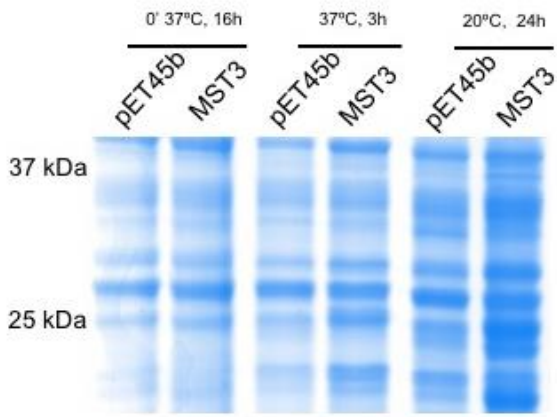

Figure 7. Optimization of MST3 Expression Conditions. We evaluated expression of MST3 (36kDa) in multiple growth conditions using SDS-PAGE. Coomassie staining indicates no significant expression of MST3. (0' denotes immediate induction with IPTG).

After much trial and error, the best expression conditions for MST3 in C41(DE3) were empirically determined to be a 24 hour incubation at 22°C following mid-log induction. With these conditions, significant autophosphorylation activity and substrate phosphorylation were observed (**Figure 8**). The presence of the his-tagged MST3 truncation was also confirmed by western blot using an anti-6xHis primary antibody (**Figure 8**).

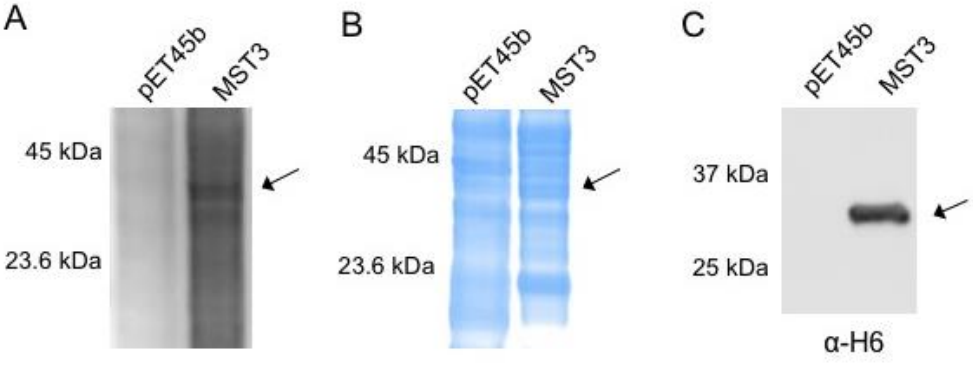

Figure 8. Expression of MST3. We evaluated expression of MST3 (36kDa) after a 24 hour incubation at 22°C. **A)** Pro-Q diamond staining reveals autophosphorylation of MST3 and substrate phosphorylation. **B)** Coomassie staining of gel in panel A for loading control. **C)** Western blot with α-6X His-antibody confirms MST3 expression. pET45b is an empty vector negative control.

Characterizing MST3 Specificity

Following the expression of MST3 as described above, the cell lysate was enriched for phosphopeptides and analyzed using LC-MS/MS. From the spectra obtained, 51 unique phosphopeptides were identified. After visualization with pLogo, a preliminary specificity motif for MST3 was generated: G-[S/T]*-Φ-Basic (* denotes the phosphoacceptor and Φ denotes any hydrophobic residue). Inspection of the unfixed pLogo (**Figure 9, panel A**) reveals a clear preference for threonine over serine as the phosphoacceptor. This is not common within the kinome, but is expected for the Ste20 family[17]. Two phosphotyrosines were identified which are most likely false positives from bioinformatic analysis or previously unidentified *E. coli* endogenous phosphorylation sites as there is no evidence to suggest MST3 or other Ste20 family members have any tyrosine kinase activity.

When the phosphoacceptor is fixed as either serine or threonine (**Figure 9, panel B and C respectively**), the glycine in the -1 position maintains its statistical significance. The fact that it is maintained even when a small number of phosphosites are evaluated (17 in the serine-centered pLogo) speaks to the importance of the -1 position to MST3 specify. No other single residue achieves statistical significance in the pLogos (fixed or unfixed), however there is a clear preference for a hydrophobic residue in the +1 position. The hydrophobic residues differ slightly between the serine and threonine centered pLogos. The serine centered pLogo particularly favors [L/V/F] residues in the +1 position whereas the threonine centered pLogo shows a preference for [L/V], but not phenylalanine in the +1 position. Similarly, in the +2 position, no single residue achieves statistical significance but there is a preference for a basic residue ([K/R] in the threonine centered pLogo and R in the serine centered pLogo). Overall, the fixed and unfixed pLogos show the same overall G-[T/S]*-Φ-Basic motif for MST3 but more sites should be identified to increase the significance of the motif. With the addition of more sites it

will also be interesting to see if the threonine and serine motifs diverge further and if the motif collapses to a single residue in the +1 and/or +2 positions.

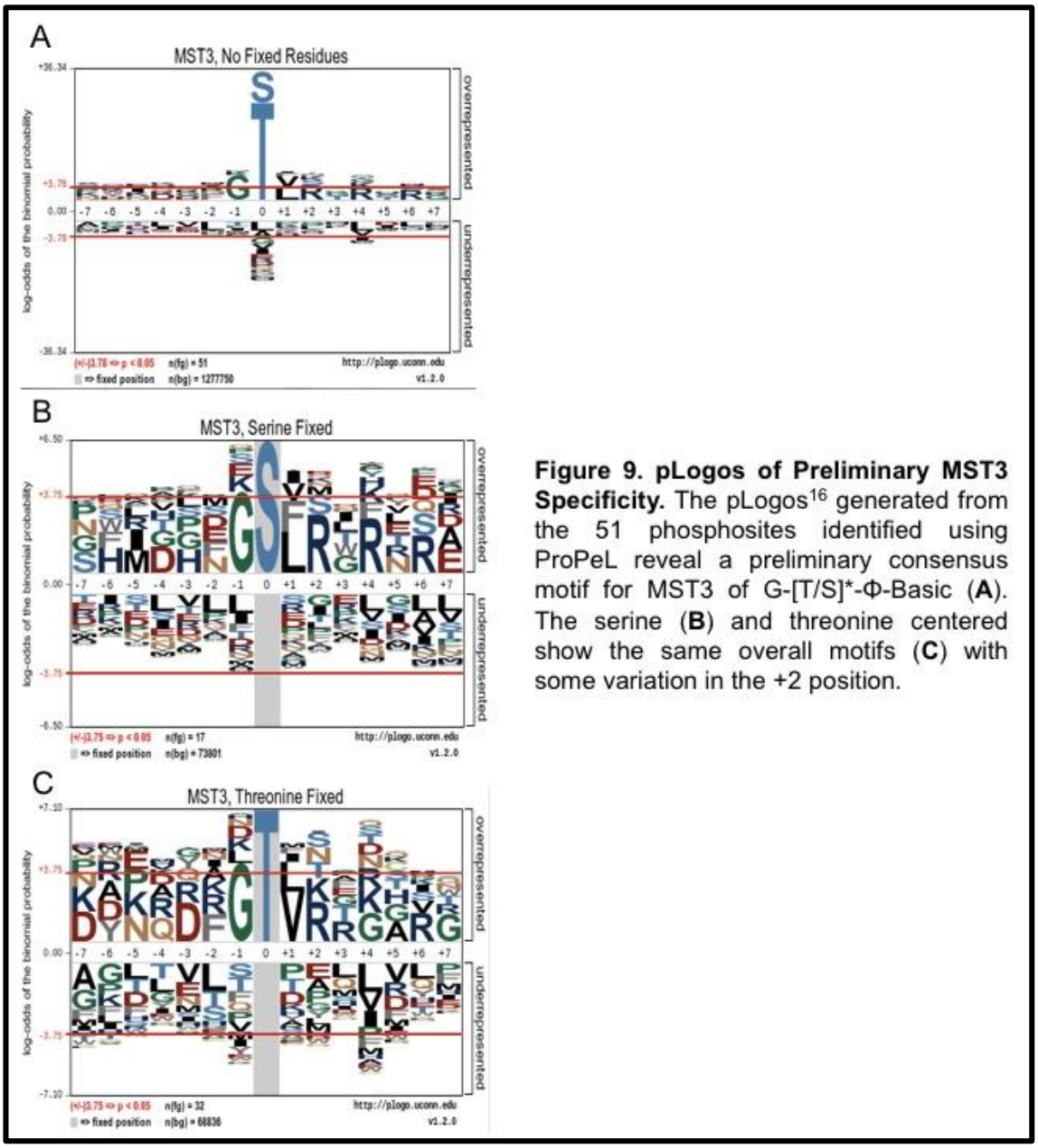

Figure 9. pLogos of Preliminary MST3 Specificity. The pLogos[16] generated from the 51 phosphosites identified using ProPeL reveal a preliminary consensus motif for MST3 of G-[T/S]*-Φ-Basic (**A**). The serine (**B**) and threonine centered show the same overall motifs (**C**) with some variation in the +2 position.

Predicting Novel MST3 Substrates

The preliminary results are sufficient to be able to predict MST3 substrates using an internal version of *scan-x*[34,35]. Only threonine sites were scored based on the strong preference of MST3 for threonine substrates. Candidate sites were scored against the position weight matrix generated from the threonine-centered MST3 pLogo[15] which is based upon the *motif-x* algorithm[36]. Putative sites were scored using *scan-x* and compared to the scores of known MST3 sites[10,17], which ranged from 12.8817 (EPS8^{T317}) to 1.36448 (Caspase-3^{T20}), scored in arbitrary units (AU) (see **Appendix** for details). This wide range of values can be attributed to the fact that we used a preliminary pLogo derived from a small foreground data set, which makes it difficult to determine thresholds. These scores are still useful for ranking purposes and a standard to which we can compare the putative sites identified.

Known MST3 interacting partners identified using the STRING Database[37] were queried for potential MST3 phosphorylation sites. Among the 30 highest scoring predicted sites, 2 are orphan phosphorylation sites (i.e. the upstream kinase remains unidentified) on the furry homolog protein (FRY). These sites are FRY^{T951} (score: 7.87595), and FRY^{T961} (score: 7.21701) (see **Appendix** for all list of all sites predicted by *scan-x*). Testing the FRY sites should be prioritized due to existing evidence in the literature that suggests that FRY may regulate dendritic arborization[38], and that MST3 regulates cytoskeletal organization of neurons especially at excitatory synapses[17–19,39]. With their physiologic roles in mind, it will be interesting to see if FRY is a true substrate of MST3. An *in vitro* [γ-^{32}P]ATP assay is one of the ways these predictions can be validated.

Conserved Herpes Virus Protein Kinases

Beginning with plasmids generously donated from the Kalejta Lab[25], all 8 full-length CHPKs

were cloned into the pET45b bacterial expression vector as 6XHis-tagged fusion protieins using directional ligation cloning[40]. Then, each was individually transformed into C41(DE3) (used for all samples unless otherwise noted) and Rosetta2 cells. C41(DE3) cells were chosen for their enhanced ability to express toxic heterologous proteins[32]. The constructs were also transformed into Rosetta2 cells which overexpress the tRNAs of seven rare codons because *E. coli* express each of the 61 tRNAs in differing amounts from eukaryotic cells and thus may need to be modified to be supplemented to better match the codon usage of the kinase transcript[41].

Expression of the kinases was attempted under different conditions. The solubility of the yield was assessed by lysing the sample in two different buffers, one denaturing condition with urea (used for all samples unless otherwise noted) and one native condition without urea. Urea disrupts the plasma membrane but also has the potential to disrupt the membrane of inclusion bodies[42]. In bacteria, misfolded proteins are often insoluble and sequestered in inclusion bodies. ProPeL will not be successful if the kinase is retained in an inclusion body, because it will not be able to interact with the majority of cellular bacterial proteins. The presence of the kinase band on a gel from urea-containing lysate (but not native conditions) indicates the kinase is being expressed but could be in an inclusion body. By lysing the cells in a urea-free buffer, which is sufficient to disrupt the *E. coli* cell wall but not the inclusion bodies, it can be determined if the kinase is both expressed and soluble.

The duration of incubation (3, 16 or 24 hours) was first varied. The 24 hour incubations were carried out at room temperature (20-22°C) rather than 37°C to prevent overgrowth of the culture media. Testing these expression conditions did not yield any expression of VZV Orf47, HHV-1 UL13, HCMV UL97 or EBV BGLF4 (**Figure 10,11**). The same conditions were also tested for HHV-7 U69 and KSHV Orf36 in both C41(DE3) cells and Rosetta2 cells (**Figure 12**). Aeration of the culture was also adjusted by changing the surface area to volume ratio by varying the

volume of the culture media which was incubated during a 3 hour incubation at 37°C following mid-log induction (**Figure 13**). The endogenous *E. coli* proteins around 50 kDa make it difficult to tell if there was expression of HHV-7 U69 or KSHV Orf36, so a western blot was performed using an anti-6xHis primary antibody which clearly revealed soluble expression of KSHV Orf36 but no expression of HHV-7 U69 in C41(DE3) cells after a 24 hour incubation at 20°C (**Figure 14**).

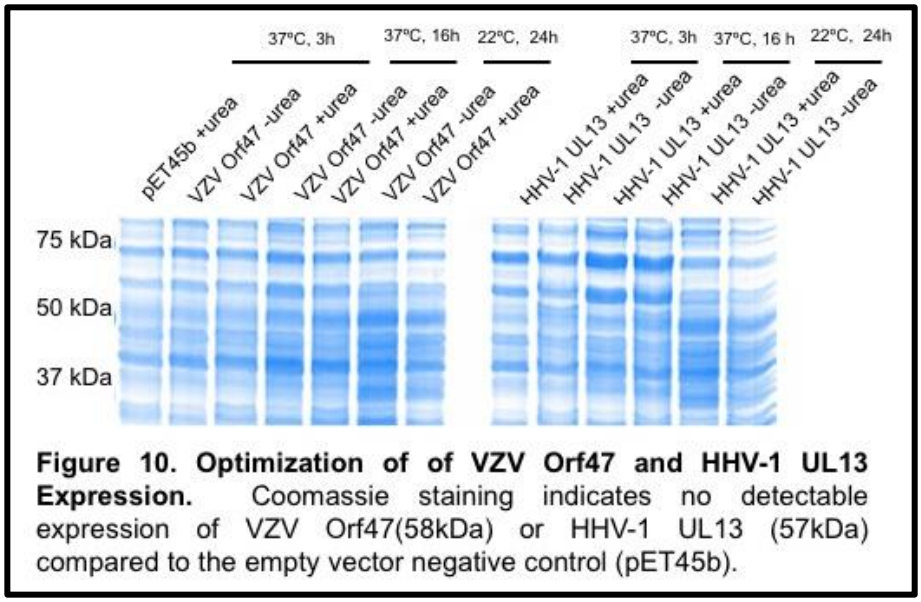

Figure 10. Optimization of of VZV Orf47 and HHV-1 UL13 Expression. Coomassie staining indicates no detectable expression of VZV Orf47(58kDa) or HHV-1 UL13 (57kDa) compared to the empty vector negative control (pET45b).

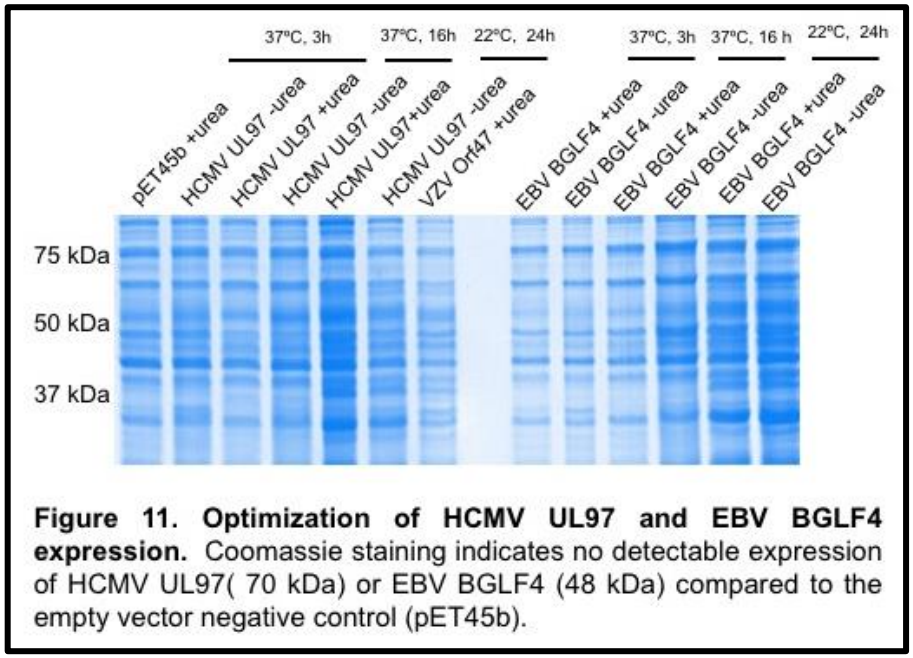

Figure 11. Optimization of HCMV UL97 and EBV BGLF4 expression. Coomassie staining indicates no detectable expression of HCMV UL97(70 kDa) or EBV BGLF4 (48 kDa) compared to the empty vector negative control (pET45b).

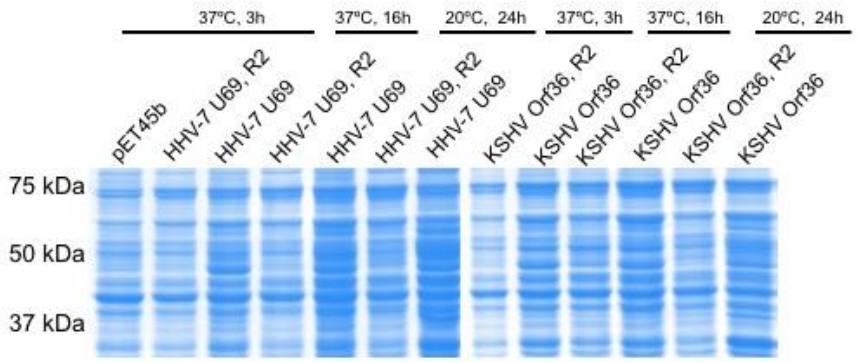

Figure 12. Optimization of HHV-7 U69 and KSHV Orf36 Expression. Coomassie staining is inconclusive regarding expression of HHV-7 U69 (53kDa) or KSHV Orf36 (50 kDa) compared to the empty vector negative control (pET45b). (R2 denotes expression in codon-optimized Rosetta2 cells).

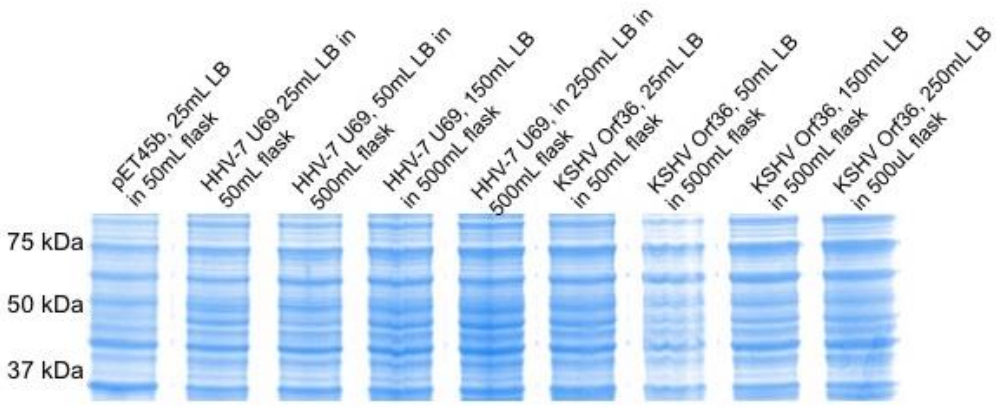

Figure 13. Optimization of HHV-7 U69 and KSHV Orf36 Expression. All cultures were induced at mid-log and incubated for 3 hours at 37°C. The ratio of volume of LB used to volume of Erlenmeyer flask was varied. Coomassie staining indicates no detectable expression of HHV-7 U69 (53kDa) or KSHV Orf36 (50 kDa) compared to the empty vector negative control (pET45b).

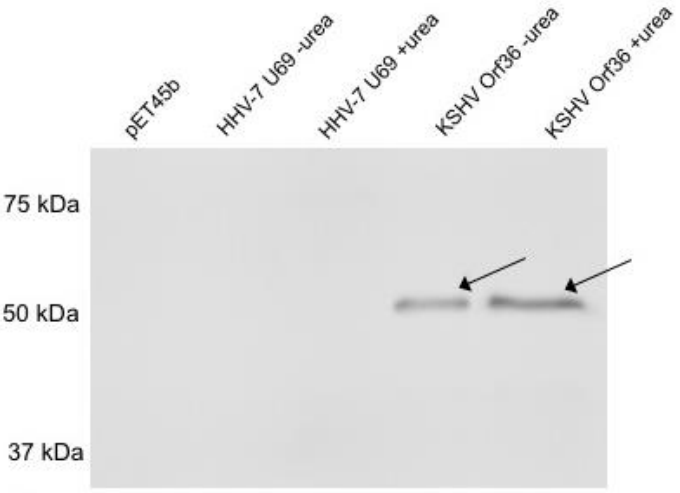

Figure 14. Expression of KSHV Orf36. Western blot with α-6X His-antibody confirms KSHV Orf36 (50kDa) expression. pET45b is an empty vector negative control.

Since first obtaining expression of KSHV Orf36, alternative conditions have been tested in the attempt to increase yield and catalytic activity. Nine different colonies were selected to test for differential expression between them using the conditions that produced soluble expression (24 hour incubation at 20°C). There were no significant differences observed between colonies (**Figure 15**). The KSHV Orf36 construct was transformed into ArcticExpress (DE3)RIL cells, which express the tRNAs of three rare codons as well as cold-adapted chaperonins. These cells are specifically designed to grow at low temperatures (10-13°C) in order to increase the soluble yield of heterologous protein[43]. However, no soluble expression of KSHV Orf36 was observed in ArcticExpress (DE3)RIL cells (**Figure 16**). This prompted experimentation with inducing expression of the kinase in a different way: auto-induction[44].

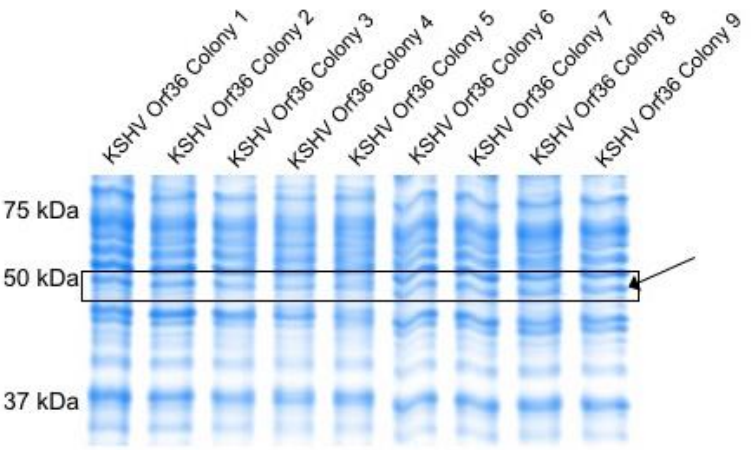

Figure 15. Optimization of KSHV Orf36 Expression. Coomassie staining shows no significant difference in KSHV Orf36 (50 kDa) expression between colonies.

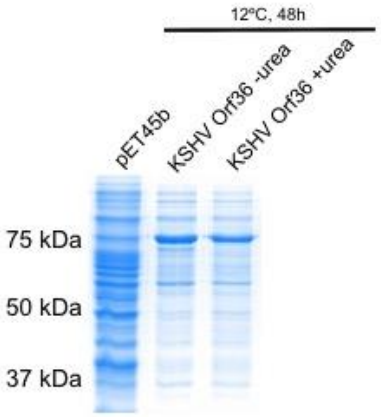

Figure 16. Optimization of KSHV Orf36 Expression. Coomassie staining indicates no detectable expression of KSHV Orf36 in Arctic Express (DE3) RIL cells compared to the empty vector negative control (pET45b).

Auto-induction does not use IPTG to induce protein expression. Rather, the growth media contains both glucose and lactose and makes use of the diauxic growth of *E. coli*[45] to control growth rate and expression of a heterologous protein under the control of the lac promoter. In the presence of glucose and lactose, the bacteria will preferentially metabolize glucose and not express the genes under the control of the lac promoter. In the absence of glucose, the bacteria will metabolize lactose, the allolactose produced as a byproduct will disinhibit transcription, transcription of the genes under the control of the lac operon will begin, and the *E. coli* grow more slowly. Auto-induction media contains only enough glucose to support bacterial growth for a few hours at 37°C at which point the bacteria will switch over to metabolize lactose[44,46]. This ensures that the heterologous protein is only expressed once the culture is robust enough to handle the metabolic demand of transcription and translation of that protein. Expression of KSHV Orf36 was evaluated using the ZYM-5052 auto-induction method described by Studier[44] in multiple conditions but did not show any significant expression (**Figure 17**).

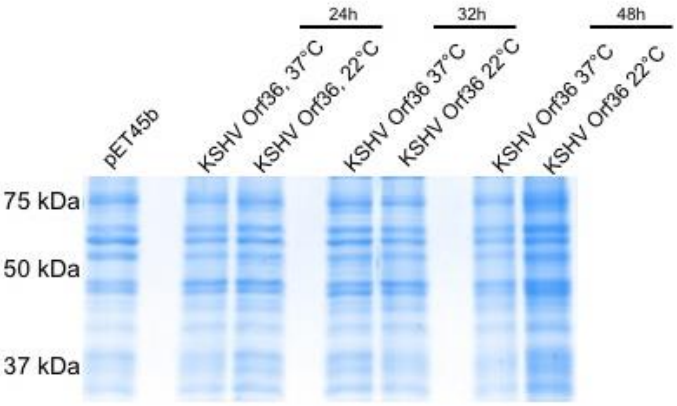

Figure 17. Optimization of KSHV Orf36 Expression. Coomassie staining indicates no detectable expression of KSHV Orf36 (50 kDa) after auto-induction compared to the empty vector negative control (pET45b).

Discussion

MST3

The correct expression conditions for MST3 were empirically determined to be a 24 hour incubation at 22°C following mid-log induction in C41(DE3) cells. Using ProPeL, a preliminary specificity motif for MST3 was obtained: G-[T/S]*-Φ-Basic with a preference for threonine as the phosphoacceptor residue which agrees published data[10,17]. This motif is based upon 51 unique phosphopeptides and therefore should be considered preliminary. In the future more sites will need to be identified to generate a more robust motif. There are some observable differences between specific residues in the +1 and +2 positions in the serine and threonine centered pLogos (**Figure 9, panel B and C**). With the identification of more sites, it will be interesting to see if these differences are sustained or the motifs will converge. Additionally, there is a trend towards a basic preference in the +4 position for both the serine and threonine centered pLogos, but especially for serine. The +4 position of the serine centered pLogo shows R/K/H residues interrupted by phenylalanine and the threonine centered pLogo shows only R/K residues. More sites will need to be identified to evaluate the significance of a basic preference

in the +4 position. It will also be interesting to see if additional sites reveal a true preference for specific amino acids in these positions, or if they collapse to electrostatic/hydrophobic interactions (i.e. if there is a real preference specifically for arginine in the serine centered +2 position, or if it is a general preference for basic residues).

From the phosphosites identified from ProPeL, *scan-x* was used to make predictions on potential threonine substrates. The query was limited to 40 proteins known to interact with MST3 identified from the STRING Database[37]. The search was restricted in this way to increase the probability that any sites *scan-x* predicted to be phosphorylated would be true MST3 phosphorylation sites. Many candidate peptides scored well within the range of scores of true positive phosphorylation site (see **Appendix**).

The identification of the putative MST3 phosphorylation sites, FRY$^{T951, T961}$, is exemplary of the utility of ProPeL combined with *scan-x.* These sites on FRY are of particular interest because they are orphan phosphorylation sites[10] and the functional role aligns with that of MST3 in that it is critical for the arborization of dendrites[47]. FRY is known to be involved in the Hippo pathway in *D. melaogaster* and MST3 is one of the orthologs of Hippo in humans therefore increasing the potential that these sites are indeed phosphorylated by MST3. However, further studies are needed to validate these predictions and have the potential to elucidate the mechanisms of neuron morphogenesis. As additional MST3 phosphorylation sites are identified and a stronger specificity motif is generated, potential targets from *scan-x* analyses will be refined and re-evaluated further in the hope of better understanding MST3's physiologic role.

Conserved Herpes Virus Protein Kinases

Though the levels of CHPK expression necessary to proceed with ProPeL have not yet been achieved, it is likely that new techniques will make it possible to do so. A cell-free protein

expression system[48] derived from human cells is a promising option in this case. In this system, the kinase would be post-transnationally modified as it is *in vivo* which may help it to fold properly and increase its stability. There is evidence to suggest this may work as CHPKs have been successfully expressed in human cell lines[24,49,50]. If any CHPK was expressed successfully in a cell-free system (or a traditional heterologous expression line such as HEK293T cells), the kinase could be purified out and ProPeL could be continued *in vitro*. This involves adding it to *E. coli* lysate and allowing it to phosphorylate *E. coli* proteins as it would in the normal ProPeL methodology.

It is also worth considering why we have experienced such difficulty in expressing CHPKs. In the case of KSHV Orf36 specifically, low levels of expression have been observed inconsistently and only in one condition. The kinase that is present is almost entirely soluble (**Figure 14**) which suggests that it is likely not a folding problem for that particular protein. The problem likely falls into one of two major categories: the kinase that is being produced by the bacteria is cytotoxic and rapidly degraded by the cell or the kinase is being expressed at very low levels to begin with. If expression of CHPKs is cytotoxic to the cell, yield may be increased by having the kinases secreted into the culture media using an endogenous mechanism such as the TAT pathway[51]. A fusion protein of the kinase and the TAT signal peptide would need to be created. Upon expression of the kinase fusion protein, it will be secreted by the bacteria, and can then be purified from the culture media and used in an *in vitro* ProPeL experiment.

It also must be determined if the CHPKs are being expressed at all. It is possible that nonsense mutations are being introduced early on during the induction. If the selection pressure favoring the mutation is strong enough due to decreased metabolic burden or the absence of a cytotoxic product, cells expressing that mutation would be heavily selected for. In this scenario it is entirely possible that early on in the incubation most of the *E. coli* population would not be

producing any functional kinase. In the future, this hypothesis could be tested by taking a sample before induction and at time points after induction and sequencing the plasmids they are carrying. If this type of selection was found to be occurring, the CHPKs could be placed under the control of a more tightly regulated promoter such as the arabinose promoter[52]. This may prevent leaky expression early in the incubation, before induction with IPTG, and thereby reduce the selection pressure for any mutations that may arise.

Materials and Methods

1. Cloning

The MST3 plasmid was purchased from Harvard Medical School's PlasmID Repository (HsCD00000386). The Kalejta Lab provided all CHPK plasmids[25]: HHV-1 UL13 (AddGene #29967), HHV-2 UL13 (not publically deposited), VZV Orf47 (AddGene #26698), EBV BGLF4 (AddGene#37936), HCMV UL97 (AddGene #26687), HHV-6 U69 (AddGene#26689), HHV-7 U69 (AddGene#26693), KSHV Orf36 (AddGene #26696). A constitutively active truncation of MST3 (amino acids 1-325) and the full length CHPKs were cloned into the ampicillin resistant bacterial expression plasmid pET45b (Novagen) by traditional restriction site ligation or overlap extension[31] and transformed into either C41(DE3) cells (Lucigen), Rosetta2 cells (Novagen) or ArcticExpress (DE3) RIL cells (Agilent Technologies). All plasmids were sequence verified after transformation by Sanger sequencing.

2. Expression

Overnight stocks were made from individual colonies. Overnight culture was added to fresh LB + appropriate antibiotic (ampicillin for all and additionally chloramphenicol for Rosetta2 and gentamycin for ArcticExpress (DE3)RIL) in a ratio of 500 µL culture to 25 mL LB and incubated in a shaking incubator at 37°C and 250 rpm until reaching mid-log (OD_{600} 0.4-0.6), then induced with 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) (some samples were induced immediately). For auto-induction, 500 µL of overnight culture was added to 25 mL of ZYM-5052 media[44] + the appropriate antibiotic. Time and temperature of incubation were varied as discussed. The cells were then pelleted by centrifugation at 6,000 g for 15 minutes at 4°C and stored at −80°C.

3. Lysis and Evaluating Success of *in vivo* Phosphorylation

Pellets were resuspended in 2 mL lysis buffer (50 mM Tris-HCl, pH 8.2, 75 mM NaCl, 1X protease and phosphatase inhibitors (Thermo Scientific), 1 mM PMSF, with or without 8 M urea)

per gram of wet pellet. Cells were lysed by sonication and clarified by centrifugation, spinning at 20,000 g for ≥ 30 minutes at 4°C. Protein concentration of the clarified lysate was measured by BCA assay (Thermo Scientific) and evaluated by SDS-PAGE, using the in-gel Pro-Q Diamond Phosphoprotein Gel Stain (Thermo Fischer Scientific) for observing phosphorylation, and GelCode Blue Safe Protein Stain (Thermo Fischer Scientific), a Coomassie type stain, for loading control. Some gels were evaluated with Coomassie exclusively.

4. Western Blotting

Western blotting for MST3, HHV-7 U69 and KSHVOrf36 used the primary antibody Anti-6xHis (NeuroMab clone N144/14, RRID: AB_10671171, UC Davis/NIH NeuroMab Facility) at 1:1000 dilution, and IRDye 800CW Goat anti-Mouse IgG secondary antibody (LI-COR Biosciences) at 1:5000 dilution.

5. Sample Preparation for LC-MS/MS

a. Methanol/Chloroform Extraction

10 mg of protein was isolated from the appropriate volume of clarified lysate prepared during step 3 by methanol/chloroform extraction. 4X sample volume of methanol was added to the lysate then 1X sample volume of chloroform then 3X sample volume of water, vortexing after each addition. The sample was centrifuged at 14,000 g for 5 minutes at 4°C. The upper aqueous layer was removed without disturbing the protein disc. 4X sample volume was added to the tube containing the protein disc and organic layer, vortexed and centrifuged again for 5 minutes at 14,000 g at 4°C. As much methanol was removed from the tube without the disturbing the protein disc as possible. The pellet was then left to air dry for about 10 minutes.

b. Protein Reduction, Alkylation and Tryptic Digest

The dry pellet from step 5a was resuspended in lysis buffer containing urea at a concentration of 10 mg/mL. Then dithiothreitol (DTT) was added to reach a final concentration of 5 mM and the sample was incubated at 56°C for 25 minutes. Once the

sample was cooled to room temperature, iodoacetamide was added to a final concentration of 14 mM and incubated at room temperature for 30 minutes in the dark. The reaction was quenched by adding an additional 5 mM DTT. After incubating at room temperature for 15 minutes in the dark, the mixture was diluted 1:5 in 25 mM Tris-HCl, pH 8.2 to reduce the urea concentration to 1.6 M. $CaCl_2$ was added to a final concentration of 1 mM before adding 100 µg of trypsin per 10 mg protein and incubated for 16 hours at 37°C. The digestion was stopped by adding triflouroacetic acid to 0.4% (v/v) or until pH<2.0 then centrifuged for 10 minutes at 2,500g at room temperature. The pellet was discarded and the peptides were stored at 4°C.

c. Peptide Desalting

Desalting was performed on 200 mg SEP-PAK Cartridges (Waters). The column was conditioned using 6 mL of acetonitrile followed by 2 mL of Elution Solution (50% acetonitrile, 0.5% acetic acid). The column was then equilibrated with 6 mL Wash Solution (0.1% triflouroacetic acid). The sample was then loaded and desalted with 6mL of Wash Solution then washed with 500 µL of 0.5% acetic acid and eluted with 1mL of Elution Solution. Each aliquot was snap frozen, dried then stored at -20ºC.

d. TiO_2 Phosphoenrichment and Terminal Desalting

The peptides were first resolubilized in 1 mL of TiO_2 Binding Solution (50% acetonitrile and 2 M lactic acid). TiO_2 beads were conditioned by washing in 50X bead volume of TiO_2 Binding Solution, centrifuging at 600 g for 30 seconds. The supernatant was removed and the wash was repeated. The beads were resuspended in TiO_2 Binding Solution to a final concentration of 10 µg/µL. TiO_2 beads were added in a 1:1 ratio of beads to peptides and additional TiO_2 Binding Solution was added to a final peptide concentration of 1 mg/mL. The peptide and bead mixture was incubated at room temperature at 1400 rpm for 1 hour. After incubation, the beads were pelleted by centrifuging at 600 g for 30 seconds. The

supernatant was removed and the beads were washed with 1 mL TiO_2 Binding Solution. The wash was repeated twice more for a total of three washes. The beads were resuspended in 200 µL TiO_2 Binding Solution. All proceeding steps were performed by centrifuging at 2,000 g for the minimum amout of time required to pass all of the liquid (about 30 seconds/50 µL of liquid). C18 Stage Tips[53] were prepared and conditioned with 50 µL of 100% methanol. They were then prepared by washing with 50 µL of Elution solution A (50% acetonitrile and 0.5% acetic acid) and equilibrated by washing twice with 50 µL of TiO_2 Binding Solution. The TiO_2 bead mixture was loaded and centrifuged. The beads were then washed twice with 150 µL TiO_2 Binding Solution. The Stage Tips were equilibrated with 100 µL of 1% formic acid and the phosphopetides were eluted with two washes of 150 µL of 50 mM KH_2PO_4, pH 10. The phosphopeptides wer washed with 100 µL of 1% formic acid and eluted off the disk with 100 µL of Elution Solution A.

6. Analysis by LC-MS/MS

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) was performed as described previously[12,28]. Each sample was resuspended in 3% Acetonitrile and 0.125% formic acid and injected, using a C18 nanocapillary column, directly into an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was run using a TOP10 method and each sample's spectra were analyzed using SEQUEST[12,54].

7. Visualization of Motifs using pLogo

Probability Logo Generator, or pLogo, is a Web tool that allows users to visualize sequence motifs in a new way[15]. The tool, based loosely upon the *motif-x* algorithm[36], compares the observed frequencies within a particular foreground data set to their expected background probability of occurrence. In the context of ProPeL, the foreground data set is the set of phosphopeptides identified by MS/MS and the background data set is the *E. coli* proteome. The result of the analysis is displayed as a pLogo (known human PKA specificity[55] is shown

as an example, **Figure 18**) with the log odds of the binomial probability on the logarithmic y-axis and the relative position to the phosphoacceptor on the x-axis. Residue heights are scaled proportional to their statistical significance and appear above or below the x-axis corresponding to overrepresentation or underrepresentation in the foreground data set, respectively. Each of the 20 amino acids is displayed at every non-fixed position, with the most significant over- and underrepresented residue nearest to the x-axis. The red horizontal line marks the threshold for statistical significance, corresponding to a Bonferroni-corrected $p = 0.05$. Unlike other sequence logos[56], pLogo is an interactive tool that allows the user to select one residue per position to "fix", which subsets the data to only compare foreground and background sequences which contain a "fixed" residue in a specific position, thereby creating a new pLogo of conditional probabilities. This feature is especially useful when comparing specificity for different phosphoacceptors (i.e. serine- vs threonine-centered motifs) or examining correlated positions within motifs. We recently demonstrated strong agreement between pLogo representations of kinase specificity and traditional [γ-^{32}P]ATP assays further validating the utility of this tool[28].

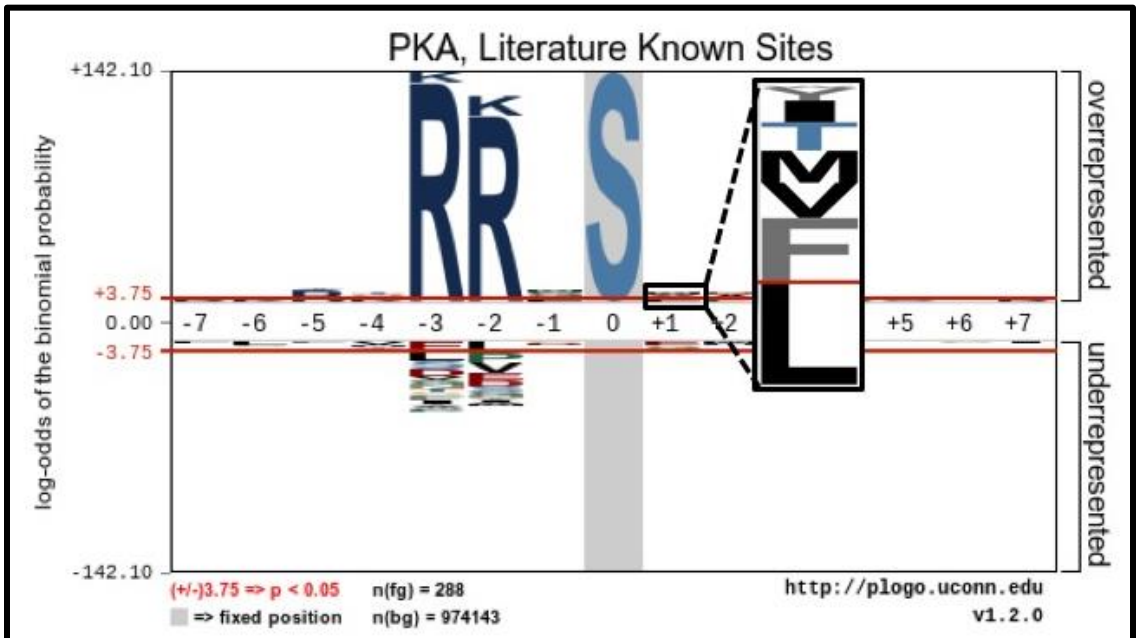

Figure 18. Example pLogo: PKA specificity. This pLogo has been created from 288 previously identified phosphorylation sites[10, 16] (*n(fg)*) and 974,143 sites from the whole *E. coli* proteome (*n(bg)*). The threshold for statistical significance has been set with a Bonferroni-corrected p-value of <0.05. With the phosphorylated serine fixed, pLogo clearly shows the RRxS*Φ (where * denotes the phosphoacceptor and Φ denotes a hydrophobic residue) motif, which agrees with the accepted PKA motif[54].

8. Phosphosite List Filtering

Prior to motif analysis, a master negative control list was generated by pooling phosphopeptides previously identified in negative control experiments[12], previously identified endogenous E. coli phosphorylation sites[13,14], and phosphorylation sites identified in empty vector and kinase dead negative control experiments. Phosphorylation sites on this master negative control list were removed from the MST3 dataset to generate a final list of kinase-specific phosphorylation sites. Peptide lists from multiple runs were merged, and redundant peptides were removed prior to motif analysis.

9. Scoring Known and Candidate Peptides

scan-x analyses of known and potential MST3 substrates were carried out using an internal version of the *scan-x* software[34]. True positive and candidate peptides were scored for a

goodness-of-fit using the MST3 position weight matrix (PWM) obtained through ProPeL. Known verified human substrates were retrieved from the PhosphoSitePlus database[10] and an earlier study[17], while candidate substrates were identified from the STRING Database[37]. The predictive capacity of *scan-x* given the PWM being used was evaluated by comparing the mean score a random set of threonine centered peptides from the human proteome to the mean score of true positives (**Figure 19**). There is a statistically significant difference between the mean scores indicating that *scan-x* can discriminate between random peptides and MST3 substrates.

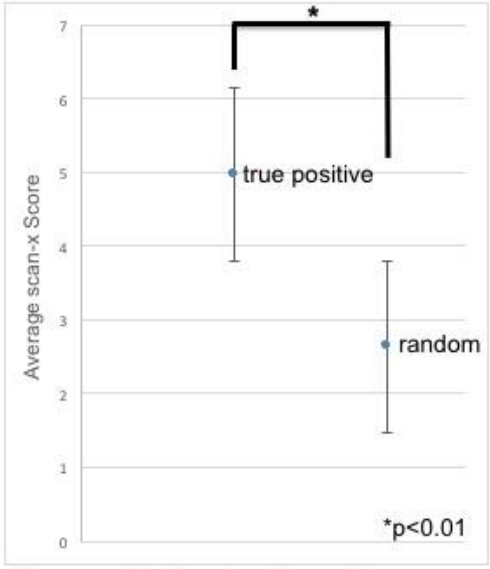

Figure 19. Evaluation of *scan-x*. Mean scores of true positives and random peptides are significantly different.

References

1. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The Protein Kinase Complement of the Human Genome. *Science* **298,** 1912–1934 (2002).
2. Hanks, S. K. & Hunter, T. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J.* **9,** 576–596 (1995).
3. Olsen, J. V. *et al.* Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell* **127,** 635–648 (2006).
4. Lahiry, P., Torkamani, A., Schork, N. J. & Hegele, R. A. Kinase mutations in human disease: interpreting genotype–phenotype relationships. *Nat. Rev. Genet.* **11,** 60–74 (2010).
5. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446,** 153–158 (2007).
6. Faux, M. C. & Scott, J. D. More on target with protein phosphorylation: conferring specificity by location. *Trends Biochem. Sci.* **21,** 312 (1996).
7. Corson, L. B., Yamanaka, Y., Lai, K.-M. V. & Rossant, J. Spatial and temporal patterns of ERK signaling during mouse embryogenesis. *Development* **130,** 4527–4537 (2003).
8. Ubersax, J. A. & Ferrell Jr, J. E. Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell Biol.* **8,** 530–541 (2007).
9. Miller, M. L. *et al.* Linear Motif Atlas for Phosphorylation-Dependent Signaling. *Sci Signal* **1,** ra2–ra2 (2008).
10. Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43,** D512–D520 (2015).
11. Hutti, J. E. *et al.* A rapid method for determining protein kinase phosphorylation specificity. *Nat. Methods* **1,** 27–29 (2004).
12. Chou, M. F. *et al.* Using Bacteria to Determine Protein Kinase Specificity and Predict Target Substrates. *PLOS ONE* **7,** e52747 (2012).
13. Macek, B. *et al.* Phosphoproteome Analysis of E. coli Reveals Evolutionary Conservation of Bacterial Ser/Thr/Tyr Phosphorylation. *Mol. Cell. Proteomics* **7,** 299–307 (2008).
14. Soares, N. C., Spät, P., Krug, K. & Macek, B. Global dynamics of the Escherichia coli proteome and phosphoproteome during growth in minimal medium. *J. Proteome Res.* **12,** 2611–2621 (2013).
15. O'Shea, J. P. *et al.* pLogo: a probabilistic approach to visualizing sequence motifs. *Nat. Methods* **10,** 1211–1212 (2013).
16. Thompson, B. J. & Sahai, E. MST kinases in development and disease. *J. Cell Biol.* **210,** 871–882 (2015).
17. Ultanir, S. K. *et al.* MST3 Kinase Phosphorylates TAO1/2 to Enable Myosin Va Function in Promoting Spine Synapse Development. *Neuron* **84,** 968–982 (2014).
18. Lorber, B., Howe, M. L., Benowitz, L. I. & Irwin, N. Mst3b, an Ste20-like kinase, regulates axon regeneration in mature CNS and PNS pathways. *Nat. Neurosci.* **12,** 1407–1414 (2009).
19. Tang, J. *et al.* Cdk5-Dependent Mst3 Phosphorylation and Activity Regulate Neuronal Migration through RhoA Inhibition. *J. Neurosci.* **34,** 7425–7436 (2014).
20. Newey, S. E., Velamoor, V., Govek, E.-E. & Van Aelst, L. Rho GTPases, dendritic structure, and mental retardation. *J. Neurobiol.* **64,** 58–74 (2005).
21. Hertz, N. T. *et al.* Chemical Genetic Approach for Kinase-Substrate Mapping by Covalent Capture of Thiophosphopeptides and Analysis by Mass Spectrometry. *Curr. Protoc. Chem. Biol.* **2,** 15–36 (2010).
22. Sharma, M. *et al.* Human Cytomegalovirus UL97 Phosphorylates the Viral Nuclear Egress Complex. *J. Virol.* **89,** 523–534 (2015).
23. Gershburg, E. & Pagano, J. S. Conserved herpesvirus protein kinases. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1784,** 203–212 (2008).
24. Hume, A. J. *et al.* Phosphorylation of Retinoblastoma Protein by Viral Protein with Cyclin-Dependent Kinase Function. *Science* **320,** 797–799 (2008).
25. Kuny, C. V., Chinchilla, K., Culbertson, M. R. & Kalejta, R. F. Cyclin-Dependent Kinase-Like Function Is Shared by the Beta- and Gamma- Subset of the Conserved Herpesvirus Protein Kinases. *PLOS Pathog* **6,** e1001092 (2010).
26. Shugar, D. Viral and Host-Cell Protein Kinases: Enticing Antiviral Targets and Relevance of Nucleoside, and Viral Thymidine, Kinases*. *Pharmacol. Ther.* **82,** 315–335 (1999).

27. Mitchell, A. *et al.* The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* **43,** D213–221 (2015).
28. Lubner, J. M. *et al.* Cushing's syndrome mutant PKAL205R exhibits altered substrate specificity. *FEBS Lett.* **591,** 459–467 (2017).
29. Nolen, B., Taylor, S. & Ghosh, G. Regulation of Protein Kinases. *Mol. Cell* **15,** 661–675 (2004).
30. Huang, C.-Y. F. *et al.* Caspase Activation of Mammalian Sterile 20-like Kinase 3 (Mst3) NUCLEAR TRANSLOCATION AND INDUCTION OF APOPTOSIS. *J. Biol. Chem.* **277,** 34367–34374 (2002).
31. Anton V. Bryksin & Ichiro Matsumura. BioTechniques - Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *BioTechniques* **48,** 463–465 (2010).
32. Miroux, B. & Walker, J. E. Over-production of Proteins inEscherichia coli: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. *J. Mol. Biol.* **260,** 289–298 (1996).
33. Weickert, M. J., Doherty, D. H., Best, E. A. & Olins, P. O. Optimization of heterologous protein production in Escherichia coli. *Curr. Opin. Biotechnol.* **7,** 494–499 (1996).
34. Schwartz, D., Chou, M. F. & Church, G. M. Predicting Protein Post-translational Modifications Using Meta-analysis of Proteome Scale Data Sets. *Mol. Cell. Proteomics* **8,** 365–379 (2009).
35. Chou, M. F. & Schwartz, D. Using the scan-x Web site to predict protein post-translational modifications. *Curr. Protoc. Bioinforma.* **Chapter 13,** Unit 13.16. (2011).
36. Schwartz, D. & Gygi, S. P. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol.* **23,** 1391–1398 (2005).
37. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43,** D447–452 (2015).
38. Shih, P.-Y., Lee, S.-P., Chen, Y.-K. & Hsueh, Y.-P. Cortactin-binding protein 2 increases microtubule stability and regulates dendritic arborization. *J Cell Sci* **127,** 3521–3534 (2014).
39. Irwin, N., Li, Y.-M., O'Toole, J. E. & Benowitz, L. I. Mst3b, a purine-sensitive Ste20-like protein kinase, regulates axon outgrowth. *Proc. Natl. Acad. Sci.* **103,** 18320–18325 (2006).
40. Hartley, J. L., Temple, G. F. & Brasch, M. A. DNA Cloning Using In Vitro Site-Specific Recombination. *Genome Res.* **10,** 1788–1795 (2000).
41. Kane, J. F. Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. *Curr. Opin. Biotechnol.* **6,** 494–500 (1995).
42. Tsumoto, K., Ejima, D., Kumagai, I. & Arakawa, T. Practical considerations in refolding proteins from inclusion bodies. *Protein Expr. Purif.* **28,** 1–8 (2003).
43. Schein, C. H. & Noteborn, M. H. M. Formation of Soluble Recombinant Proteins in Escherichia Coli is Favored by Lower Growth Temperature. *Nat. Biotechnol.* **6,** 291–294 (1988).
44. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 (2005).
45. Inada, T., Kimata, K. & Aiba, H. Mechanism responsible for glucose-lactose diauxie in Escherichia coli: challenge to the cAMP model. *Genes Cells Devoted Mol. Cell. Mech.* **1,** 293–301 (1996).
46. Neubauer, P., Hofmann, K., Holst, O., Mattiasson, B. & Kruschke, P. Maximizing the expression of a recombinant gene in Escherichia coli by manipulation of induction time using lactose as inducer. *Appl. Microbiol. Biotechnol.* **36,** 739–744 (1992).
47. Emoto, K., Parrish, J. Z., Jan, L. Y. & Jan, Y.-N. The tumour suppressor Hippo acts with the NDR kinases in dendritic tiling and maintenance. *Nature* **443,** 210–213 (2006).
48. Mikami, S., Kobayashi, T., Masutani, M., Yokoyama, S. & Imataka, H. A human cell-derived in vitro coupled transcription/translation system optimized for production of recombinant proteins. *Protein Expr. Purif.* **62,** 190–198 (2008).
49. Kawaguchi, Y. *et al.* Conserved Protein Kinases Encoded by Herpesviruses and Cellular Protein Kinase cdc2 Target the Same Phosphorylation Site in Eukaryotic Elongation Factor 1δ. *J. Virol.* **77,** 2359–2368 (2003).
50. Hamza, M. S. *et al.* ORF36 Protein Kinase of Kaposi's Sarcoma Herpesvirus Activates the c-Jun N-terminal Kinase Signaling Pathway. *J. Biol. Chem.* **279,** 38325–38330 (2004).
51. Palmer, T. & Berks, B. C. Moving folded proteins across the bacterial cell membrane. *Microbiology* **149,** 547–556 (2003).
52. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177,** 4121–4130 (1995).

53. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. *Anal. Chem.* **75,** 663–670 (2003).
54. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5,** 976–989 (1994).
55. Kemp, B. E., Graves, D. J., Benjamini, E. & Krebs, E. G. Role of multiple basic residues in determining the substrate specificity of cyclic AMP-dependent protein kinase. *J. Biol. Chem.* **252,** 4888–4894 (1977).
56. Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18,** 6097–6100 (1990).